TRAITÉ DU CALCUL DES PROBABILITÉS ET DE SES APPLICATIONS

Par Émile BOREL

Professeur de Calcul des Probabilités et de Physique mathématique
à la Faculté des Sciences de Paris,
Membre de l'Institut

TOME I. — *Les principes de la Théorie des probabilités.*

1. Principes et formules classiques, par Émile Borel, rédigé par R. Lagrange.
2. Erreurs et moindres carrés, par Robert Deltheil.
3. Recherches théoriques modernes, par Émile Borel.
4. Les principes de la statistique mathématique, par C.-E. Traynard.

TOME II. — *Les applications de la Théorie des probabilités aux sciences mathématiques et aux sciences physiques.*

1. Applications à l'arithmétique et à la théorie des fonctions, par Émile Borel.
2. Probabilités géométriques, par Robert Deltheil.
3. Mécanique statistique classique, par Émile Borel, rédigé par Francis Perrin.
4. Applications à l'astronomie, par C.-V.-L. Charlier.
5. Applications aux théories physiques actuelles, par Émile Borel et Francis Perrin.

TOME III. — *Les applications de la Théorie des probabilités aux sciences économiques et aux sciences biologiques.*

1. Assurances sur la vie. Calcul des primes, par Henri Galbrun.
2. Assurances sur la vie. Calcul des réserves, par Henri Galbrun.
3. Applications à la biologie. Variations discontinues et Mendélisme, par L. Blaringhem.
4. Applications à la biologie. Variations continues et biométrique, par L. Blaringhem et C.-E. Traynard.

TOME IV. — *Applications diverses et conclusion.*

1. Applications au tir, par J. Haag.
2. Applications aux jeux de hasard, par Émile Borel.
3. Compléments divers.
4. Conclusion : la portée philosophique de la théorie des probabilités, par Émile Borel.

LES PRINCIPES

DE LA

THÉORIE DES PROBABILITÉS

PRINCIPES ET FORMULES CLASSIQUES

TRAITÉ du CALCUL des PROBABILITÉS et de ses APPLICATIONS

Par Émile BOREL

AVEC LA COLLABORATION DE L. BLARINGHEM, C. V. L. CHARLIER, R. DELTHEIL, H. GALBRUN, J. HAAG, R. LAGRANGE, F. PERRIN, P. TRAYNARD

TOME I

LES PRINCIPES DE LA THÉORIE DES PROBABILITÉS

FASCICULE I

PRINCIPES ET FORMULES CLASSIQUES

DU

CALCUL DES PROBABILITÉS

Leçons professées à la Faculté des Sciences de Paris

Par M. Émile BOREL

RÉDIGÉES

Par René LAGRANGE

Maître de Conférences à la Faculté des Sciences de Rennes

PARIS

GAUTHIER-VILLARS ET C^ie^, ÉDITEURS

LIBRAIRES DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE

55, Quai des Grands Augustins, 55

PRÉFACE.

Il y a plus d'un siècle que Laplace a publié sa *Théorie analytique des probabilités* et a fourni ainsi un modèle à tous ceux qui ont abordé ce sujet. Dans le célèbre *Essai philosophique* qui sert d'Introduction à son ouvrage, le célèbre géomètre insiste sur l'importance de la Théorie des probabilités dans les domaines divers des connaissances humaines. Depuis qu'il a écrit, cette importance n'a cessé de s'accroître; c'est grâce à la théorie des probabilités, que l'on appelle le plus souvent *Calcul des probabilités*, que les physiciens modernes expliquent les propriétés les plus cachées de l'énergie et de la matière, que les biologistes arrivent à pénétrer les lois secrètes de l'hérédité, permettant ainsi aux agronomes de sélectionner des races d'animaux et de plantes; les assurances, la prévoyance sous toutes ses formes, utilisent constamment le calcul des probabilités; les astronomes cherchent à découvrir par lui les mystères de l'univers stellaire et les mathématiciens l'emploient pour étudier les propriétés des nombres et des fonctions; le calcul des probabilités intéresse les artilleurs; il intéresse aussi, non seulement les joueurs de cartes ou de dés, qui furent ses parrains, mais tous les hommes d'action, chefs d'industries ou chefs d'armées dont le succès dépend de décisions subordonnées elles-mêmes à des données dont les unes sont connues ou calculables tandis que les autres sont incertaines et problématiques; il intéresse le statisticien, le démographe, le philosophe.

Il m'a semblé que le moment était venu de chercher à rassembler en un Traité les résultats essentiels acquis à la science dans le domaine du Calcul des probabilités et de ses applications diverses. Je ne me

dissimule pas les difficultés de cette tâche et je n'ose espérer éviter les lacunes ni même les erreurs; je m'estimerai heureux si ce Traité est suivi d'autres Traités analogues, ne présentant pas les mêmes imperfections.

Quelles que puissent être ces imperfections, j'ose espérer qu'il n'aura pas été inutile de rassembler ainsi des matières le plus souvent dispersées dans des ouvrages qui s'adressent à des publics différents. Malgré leur diversité, en effet, toutes les applications du Calcul des probabilités présentent entre elles des analogies étroites, ceux qui s'intéressent spécialement à certaines de ces applications ont intérêt à étudier les méthodes employées dans d'autres applications; tous doivent connaître les recherches théoriques et, de leur côté, les théoriciens ne doivent pas ignorer les applications qui ont été ou pourront être faites de leurs travaux. En un mot, il convient de mettre en évidence, par la publication d'un Traité, l'unité et l'importance du Calcul des probabilités, comme on l'a fait depuis longtemps pour d'autres disciplines et notamment pour la Mécanique où les recherches théoriques et les applications pratiques se sont prêtées au mutuel appui.

Une telle exposition d'ensemble d'une branche de la Science doit précéder et non suivre l'étude philosophique des principes de cette science. Le savant est comme l'homme d'action; celui-ci agit comme si le monde extérieur existe et celui-là comme si les principes de la science sont légitimes; leurs conquêtes faites, il se trouve toujours un économiste, un historien ou un métaphysicien pour les justifier. Le savant est en tout cas assuré de ne pas avoir transgressé, dans son goût de l'action, les lois de la morale, même s'il a enfreint les lois de la logique.

Mais autant la critique *a priori* risque d'être stérile, autant la critique *a posteriori* peut être féconde. Lorsqu'une science a prouvé sa vitalité par ses résultats, il vaut la peine de faire un retour en arrière et de soumettre à un examen scrupuleux les principes que l'on avait admis sans discussion. Nous sommes en effet désormais assurés que cet examen ne peut pas nous égarer au point de nous faire renoncer aux conséquences positives et pratiques, qui restent acquises en tout

état de cause; quelle que puisse être l'opinion des physiciens ou des métaphysiciens sur la réalité des fluides électriques, l'electrification de nos réseaux de chemins de fer n'en sera pas retardée d'une journée; de même, les discussions philosophiques sur les probabilités ne sauraient avoir d'action sur les calculs de nos actuaires. On ne peut d'autre part oublier l'exemple fameux de la géométrie non euclidienne · les recherches de Bolyai, de Lobatchefsky, de Riemann sur les principes de la Géométrie ont permis à Poincaré la découverte des fonctions fuchsiennes et ont fourni à Einstein le support mathématique indispensable à la théorie de la relativité De même, l'étude abstraite des principes de l'arithmétique n'a pas été sans influence sur le développement de la théorie des nombres transfinis, theorie qui a facilité le progrès de la théorie des ensembles et de la théorie des fonctions Dans ces divers cas l'étude des principes d'une science a eu comme conséquence la création de nouvelles sciences, ou du moins de chapitres nouveaux dans des sciences connexes, cette étude est donc justifiée aux yeux même du savant, indépendamment de la contribution qu'elle peut apporter à la théorie de la connaissance, qui intéresse le philosophe.

La théorie des probabilités nous conduit, dans le domaine des sciences physiques, à substituer aux explications mécaniques de l'Univers des explications statistiques; de même, dans le domaine juridique et économique, les préoccupations collectives tendent de plus en plus à primer les préoccupations individualistes

Il ne sera pas inutile, après avoir étudié les applications diverses des probabilités, de réfléchir sur l'influence que peut exercer le developpement de ces applications sur notre conception de la vérité scientifique et sur nos jugements de valeur.

Afin de réduire la période de publication de ce Traité à un petit nombre d'années, il m'a paru nécessaire de faire appel à des collaborateurs, dont on a pu lire les noms ci-dessus. Les uns sont déjà des savants éminents, qui ont apporté une contribution personnelle importante aux sujets dont ils ont bien voulu se charger; d'autres, plus jeunes, n'ont pas encore acquis la notoriété de leurs aînés; ils ne tarderont pas à être de pair avec les meilleurs, s'ils justifient les

espérances suscitées par leurs premiers travaux. Aux uns comme aux autres, j'adresse ici l'expression de ma gratitude pour leur précieuse collaboration, sans laquelle je ne me serais pas senti la force d'entreprendre cette œuvre de longue haleine. Qu'il me soit permis de remercier également la maison Gauthier-Villars pour avoir accueilli ce Traité et avoir apporté à l'exécution matérielle ses soins traditionnels.

ÉMILE BOREL.

Octobre 1924.

INTRODUCTION.

Ce premier fascicule a été rédigé par M. René Lagrange, d'après un Cours que j'ai fait à la Faculté des Sciences de Paris en 1922. Je tiens à remercier M. René Lagrange des soins qu'il a apportés à cette rédaction, qu'il a su mener à bien, tout en poursuivant de son côté d'importantes recherches personnelles. La rédaction des Notes qui terminent ce volume est entièrement due à M. René Lagrange.

PRINCIPES

ET

FORMULES CLASSIQUES

DU CALCUL DES PROBABILITÉS

CHAPITRE I.

GÉNÉRALITÉS

1. *Généralités.* — L'origine du calcul des probabilités, comme celle des autres branches des Mathématiques, se trouve dans des observations concrètes ; ce sont en effet les jeux de hasard qui lui ont donné naissance, et, d'ailleurs, ce seront encore des schémas concrets, tirés de ces jeux, qui nous permettront de rendre le calcul des probabilités plus intuitif. Les schémas les plus employés sont un dé, dont les six faces sont supposées avoir la même probabilité de chute, ou une urne contenant des boules qui ont la même probabilité d'être tirées.

Nous commencerons par un exposé purement abstrait, et, ensuite, dans les applications, nous rattacherons la réalité aux prémisses que nous nous serons données. Il résulte d'ailleurs de la grande variété de ces applications que les hypothèses que l'on peut faire sont de natures très diverses.

Nous considérons des « événements », en attachant, à ce mot, l'unique qualité d'être susceptibles de se produire, ou de ne pas se produire. Dans le premier cas, l'événement est dit « favorable » ; dans le deuxième cas, il est dit « défavorable ».

En outre, à chaque événement, nous adjoignons un nombre p, compris entre 0 et 1.

La valeur limite $p = 0$ correspond à un événement qui ne se produit jamais; au contraire, $p = 1$ exprime que l'événement se produit toujours. Ces valeurs limites correspondent à une « certitude » [1]. Si p est compris entre 0 et 1, on dit que la probabilité pour que l'événement se produise est p, et, celle pour qu'il ne se produise pas, $1 - p$; cette dernière circonstance peut être envisagée comme « l'événement contraire »; c'est ainsi que $p = 0$ exprime la certitude d'obtenir l'événement contraire.

Il peut arriver que, parmi un certain nombre d'événements distincts, il ne puisse jamais s'en produire plusieurs simultanément. C'est à cette circonstance que se rapporte l'axiome général, que l'on appelle *Principe de l'addition des probabilités*, ou *principe des probabilités totales* :

Étant donnés des événements $E_1, E_2, \ldots, E_n$, *de probabilités* $p_1, p_2, \ldots, p_n$, *s'excluant mutuellement, la probabilité pour que l'un d'eux se produise est* $p_1 + p_2 + \ldots + p_n$

En particulier, si l'un de ces événements se produit sûrement, ceci s'exprime par $p_1 + p_2 + \ldots + p_n = 1$.

La définition de la probabilité se complète par la propriété qu'exprime un deuxième principe fondamental, le *Principe de la probabilité composée*, qui se rapporte à plusieurs événements E_1, $E_2, \ldots, E_n$, considérés comme consécutifs (autrement dit, dépendant d'une variable indépendante, le temps par exemple). Voici son énoncé :

Si p_1 *est la probabilité d'un événement* E_1, *et* p_2, *la probabilité d'un événement* E_2 *quand* E_1 *s'est produit, la probabilité pour que se produise la succession* $E_1 E_2$ *est* $p_1 \times p_2$.

Il en résulte immédiatement que, pour n événements ($n > 2$), si p_i est la probabilité de E_i, supposée réalisée la suite d'événements $E_1 E_2 \ldots E_{i-1}$ ($i = 1, 2, \ldots, n$), la probabilité pour que se produise la succession $E_1 E_2 \ldots E_n$ est $p_1 p_2 \ldots p_n$.

L'application de ce principe est d'une pratique aisée dès que les

(1) Nous verrons que ces valeurs limites peuvent ne plus avoir cette signification si le nombre des cas distincts possibles est infini.

événements sont indépendants entre eux; leur ordre n'intervient plus, et p_i est alors, simplement, la probabilité de E_i.

2. Nous commencerons par étudier le cas où les probabilités sont des *nombres rationnels*; c'est ce que l'on appelle le problème des *probabilités discontinues* (ou mieux, *discontinues et finies*). On peut donner de ce problème un schéma discontinu et fini, à savoir une urne, qui nous permettra souvent de simplifier certains raisonnements.

Le nombre $p = \frac{a}{N}$ peut être en effet considéré comme représentant la probabilité d'extraction d'une boule blanche d'une urne qui contient N boules, dont a, et a seulement, sont blanches. Pour avoir le droit d'adopter cette représentation, il faut évidemment vérifier que ce schéma satisfait aux deux axiomes fondamentaux.

Si, parmi les $N - a$ boules non blanches de l'urne, b sont rouges, on a la probabilité $p' = \frac{b}{N}$ de tirer une boule rouge, et la probabilité $\frac{a+b}{N}$ de tirer une boule blanche ou rouge, ce qui vérifie bien le principe d'addition.

Supposons maintenant que l'urne contienne N sacs, contenant chacun N′ boules; et que, dans a de ces sacs, a' des N′ boules soient blanches. La probabilité de retirer un de ces a sacs est $p_1 = \frac{a}{N}$; et, un tel sac étant tiré, la probabilité d'en extraire une boule blanche est $p_2 = \frac{a'}{N'}$. D'autre part, on peut mélanger toutes les boules de l'urne; la probabilité d'obtenir une boule blanche est alors $\frac{aa'}{NN'} = p_1 p_2$; on vérifie ainsi le principe de la probabilité composée, à condition d'admettre, ce qui paraît assez naturel du fait que les sacs sont identiques entre eux et renferment tous autant de boules, qu'il revient au même de retirer directement une boule blanche, sans retirer d'abord de l'urne un des N sacs.

Remarquons que, au point de vue expérimental, on peut se borner à ne considérer que des probabilités rationnelles, car on est toujours obligé de ne conserver qu'un certain nombre de décimales; on peut, également, ne considérer que des nombres N finis, ce qui revient, en général, à négliger des probabilités infiniment petites. Cependant, il n'y a pas nécessairement avantage à adopter ce point de vue, de même

qu'il est souvent plus commode de considérer un nombre irrationnel, défini simplement, qu'un nombre rationnel approché.

3. Il existe deux autres sortes de probabilités, les *probabilités continues*, et les *probabilités dénombrables*.

Les probabilités continues peuvent être définies par un schéma géométrique. On considère un certain nombre de variables, par exemple x, y, z, et une fonction de ces variables $\varphi(x, y, z)$; ces trois variables représentent un point en coordonnées cartésiennes, et, par définition, la probabilité pour que ce point soit compris dans le parallélépipède x, $x+dx$; y, $y+dy$, z, $z+dz$ est

$$\varphi(x, y, z)\,dx\,dy\,dz.$$

Le principe des probabilités totales conduit à exprimer la probabilité, pour que ce point soit dans un domaine donné D, par l'intégrale triple

$$\int\int\int_{D}\varphi(x, y, z)\,dx\,dy\,dz.$$

Nous l'écrirons encore

$$\int_{D}\varphi(x, y, z)\,dx\,dy\,dz \quad \text{ou même, plus simplement,} \quad \int_{D}\varphi(x, y, z)\,d\omega.$$

Par suite, la fonction φ doit être telle que cette intégrale, étendue à tout le domaine que peut parcourir le point, ait la valeur 1.

C'est l'application du principe de la probabilité composée qui nous permet d'écrire la probabilité pour que deux points (x, y, z) et (x', y', z') soient, l'un dans un certain domaine V, l'autre dans un domaine V'. Par exemple, si ces deux points sont indépendants, et ont séparément les probabilités élémentaires $\varphi(x, y, z)\,dx\,dy\,dz$ et $\psi(x, y, z)\,dx\,dy\,dz$ de se trouver dans le parallélépipède infinitésimal de volume $dx\,dy\,dz$, cette probabilité composée s'exprime par l'intégrale sextuple

$$\int_{V}\int_{V'}\varphi(x, y, z)\,\psi(x', y', z')\,d\omega\,d\omega'.$$

A côté des deux espèces de probabilités que nous venons de définir, on est conduit à considérer les *probabilités dénombrables*, relatives à une infinité dénombrable d'événements possibles $E_1 E_2 \ldots E_n \ldots$, sur

lesquels on effectue une quantité dénombrable d'expériences; la série de leurs probabilités $p_1 + p_2 + \dots + p_n + \dots$ est supposée avoir une somme égale à l'unité.

4. Nous allons appliquer ces définitions et ces schémas à la résolution de certains problèmes de probabilités. Nous distinguerons les problèmes du premier et du deuxième ordre, ceux-ci étant des problèmes dont les événements sont des probabilités du premier ordre. Un exemple fera immédiatement comprendre cette distinction

Dans le jeu de pile ou face, considérons une partie de 6 coups; les regles du jeu que l'on a adoptées conduisent à certaines probabilites d'avoir un gain déterminé La recherche de ces probabilités est un problème du premier ordre. Si l'on considere alors une série de parties de 6 coups, la recherche de la répartition des gains et des pertes à la fin de cette série sera un problème du deuxième ordre.

CHAPITRE II.

PROBLÈMES DU PREMIER ORDRE.

I — PROBABILITÉS DISCONTINUES

La solution des problèmes de probabilités discontinues du premier ordre peut s'obtenir, en principe, par une simple énumération arithmétique, les seules difficultés sont d'ordre pratique.

L'étude de ces problèmes ne peut donc consister qu'en la résolution d'un certain nombre d'exemples; c'est ce que nous ferons dans ce Chapitre.

1. Problème. — *Étant données 2 urnes, contenant chacune 3 boules, numérotées* 1, 2, 3, *quelle est la probabilité pour que, une boule étant tirée de chaque urne, le numéro le plus élevé soit un* 2?

L'énumération de tous les cas possibles conduit immédiatement au résultat $p = \frac{1}{3}$.

Cependant, dans cet exemple si simple, il faut faire attention à la dépendance des événements; c'est ainsi que le n° 1, tiré de la deuxième urne, ne peut donner un cas favorable que si le numéro tiré de la première urne est le n° 2. C'est cette dépendance qui rend souvent délicate, et même compliquée, l'application du principe de la probabilité composée. C'est ce que nous allons voir dans l'étude du problème généralisé :

On considère p urnes contenant chacune n boules numérotées 1, 2, ..., n. *Une boule étant tirée de chaque urne, quelle est la probabilité pour que le numéro le plus élevé soit un nombre m déterminé* $(m \leqq n)$?

Comme nous venons de le faire remarquer, les événements favorables ne sont pas indépendants entre eux; mais il n'en est plus de même des événements contraires; en effet, pour que certains numéros ne sortent pas, il est indispensable qu'ils ne sortent dans aucun des n tirages Le principe de la probabilité composée donne alors immédiatement la probabilité $\left(\frac{n-1}{n}\right)^p$ de ne pas tirer un numéro donné, et la probabilité $\left(\frac{n-q}{n}\right)^p$ de ne pas tirer l'un quelconque de q numéros données.

Désignons par P_m la probabilité cherchée. $P_1+P_2+\ldots+P_m$ représente la probabilité pour que ne sorte aucun des numéros $m+1$, $m+2,\ldots, n$.

On obtient ainsi n relations

$$
\begin{aligned}
P_1+P_2+\ldots+P_{n-1}+P_n &= 1,\\
P_1+P_2+\ldots+P_{n-1} &= \left(\frac{n-1}{n}\right)^p,\\
P_1+P_2+\ldots+P_{n-2} &= \left(\frac{n-2}{n}\right)^p,\\
\ldots\ldots\ldots\ldots & \ldots\ldots,\\
P_1 &= \left(\frac{1}{n}\right)^p,
\end{aligned}
$$

qui, par différence, fournissent les P_m pour les différentes valeurs de m :

$$P_m = \frac{m^p-(m-1)^p}{n^p}.$$

Dans le cas particulier de 2 urnes, les numérateurs de ces probabilités sont les nombres impairs successifs

$$P_m = \frac{2m-1}{n^2}.$$

Le problème relatif à $n=10$, avec les numéros 0, 1, 2,..., 9, est d'application intéressante, puisqu'il conduit à la formation des nombres décimaux de p chiffres. Si p devient infini, il est très difficile d'en déduire des renseignements sur les nombres décimaux à une infinité de chiffres; cela tient à ce que les probabilités P_m tendent vers zéro, sans que ceci signifie l'impossibilité de l'événement correspondant. C'est ainsi que l'on ne sait rien sur la répartition des chiffres des

nombres décimaux non périodiques, et cette répartition nous apparaît comme l'effet du hasard.

2. Problème. — Dans l'étude de certains problèmes, l'analyse combinatoire peut intervenir efficacement. En voici un exemple :

On considère un événement P *de probabilité* p, *et l'événement contraire* Q. *Quelle est la probabilité d'obtenir, sur* n *expériences,* m *fois l'événement* P, *et* $n-m$ *fois l'événement* Q?

Il est clair que la probabilité d'un tirage favorable, d'arrangement donné, est $p^m q^{n-m}$, il faut donc multiplier ce nombre par le nombre d'arrangements distincts que l'on peut former avec la combinaison de m événements P et $(n-m)$ événements Q, c'est-à-dire

$$C_n^m = \frac{n!}{m!(n-m)!}.$$

En se plaçant à ce point de vue, la formule du binome, développée en respectant l'ordre des facteurs, fournit tous ces arrangements; cela résulte de sa définition. Par exemple,

$$(p+q)^2 = pp + pq + qp + qq.$$

Chaque terme est la probabilité de l'arrangement qu'il représente, de sorte que le terme $P_m = \frac{n!}{m!(n-m)!} p^m q^{n-m}$, qui correspond à une même combinaison du développement de $(p+q)^n$, est égal à la probabilité de cette combinaison.

Cette formule permet de déterminer l'événement le plus probable. Il suffit d'étudier la croissance de P_m avec m.

Formons, pour cela, le rapport

$$\frac{P_{m+1}}{P_m} = \frac{p}{q}\frac{n-m}{m+1}.$$

Il nous montre que P_m croît avec m tant que m est inférieur ou égal à $np-q$, et, au contraire, décroît pour $m > np - q$. Si $np-q$ est un entier k,

$$P_{k-1} < P_k = P_{k+1} > P_{k+2};$$

il y a deux probabilités maxima, P_{np-q} et P_{np+p}.

Si $np-q$ n'est pas entier, désignons par $k-1$ sa partie entière;

il résulte alors des inégalités

$$P_{k-1} < P_k > P_{k+1}$$

qu'il y a, dans ce cas, une seule probabilité maximum, P_k.

3. JEU DE DÉS. — Étant donnés 2 dés, la méthode d'énumération permet de former toutes les sommes que l'on peut obtenir avec eux : 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, ces sommes pouvant être amenées respectivement de 1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1 façons différentes. On déduit immédiatement, de ces résultats, la probabilité d'obtenir une somme déterminée.

L'emploi de cette méthode devient malaisé dès qu'il s'agit d'un nombre plus grand de dés, cependant, le raisonnement supporte quelques simplifications basées sur des raisons de symétrie, pour 3 ou 4 dés.

Quoi qu'il en soit, on est conduit à une méthode plus générale par la considération du produit

$$(t_1 + t_1^2 + t_1^3 + \ldots + t_1^6)(t_2 + t_2^2 + t_2^3 + \ldots + t_2^6)$$

En effet, la formation du terme $t_1^\alpha t_2^\beta \left(1 \leqq \begin{matrix}\alpha\\ \beta\end{matrix} \leqq 6\right)$ revient à amener le chiffre α sur le premier dé, et β sur le deuxième dé ; de sorte que, en identifiant t_1 et t_2 à la même lettre t, le coefficient de $t^{\alpha+\beta}$ dans le développement de $(t + t^2 + t^3 + \ldots + t^6)^2$ est le nombre de manières d'obtenir la somme $k = \alpha + \beta$, avec 2 dés.

Il est clair que ce raisonnement est valable pour un nombre quelconque de dés. On obtient ainsi le résultat général suivant :

« La probabilité d'amener une somme donnée k $(n \leqq k \leqq 6n)$, avec n dés, est le coefficient $P_k^{(n)}$ de t^k dans le développement de

$$\frac{1}{6^n}(t + t^2 + \ldots + t^6)^n \text{ »}$$

La résolution algébrique ne présente aucune difficulté. L'expression obtenue s'écrit en effet

$$\frac{t^n}{6^n}(1 - t^6)^n(1 - t)^{-n} = \frac{t^n}{6^n}\left[1 - nt^6 + \frac{n(n-1)}{2}t^{12} + \ldots + (-1)^\alpha C_n^\alpha t^{6\alpha} + \ldots\right]$$
$$\times\left[1 + nt + \frac{n(n+1)}{2}t^2 + \right.$$
$$\left. + \frac{n(n+1)\ldots(n+\beta-1)}{\beta!}t^\beta + \ldots\right].$$

On trouve, en particulier,

$$P_n^{(n)} = \frac{1}{6^n}, \qquad P_{n+1}^{(n)} = \frac{n}{6^n}, \qquad \ldots,$$

$$P_{n+5}^n = \frac{n(n+1)(n+2)(n+3)(n+4)}{5!\,6^n},$$

$$P_{n+6}^n = \frac{1}{6^n}\left(\frac{n(n+1)\ldots(n+5)}{6!} - n\right)$$

4. Problème du scrutin [1]. — *Deux candidats* A *et* B *ont obtenu respectivement* m *et* n *voix* ($m > n$). *Quelle est la probabilité pour que, dans le dépouillement,* A *ait constamment la majorité ?*

De même que dans le problème que nous avons résolu, relatif aux urnes, il est ici plus facile d'exprimer qu'une combinaison n'est pas favorable que d'exprimer qu'elle est favorable. Il est clair, en effet, qu'il suffit, pour qu'une combinaison soit négative, que A perde la majorité avec un bulletin de rang quelconque. De telles combinaisons sont de deux sortes

1° Celles où A possède, au début, la majorité; il cesse alors de l'avoir, par égalité des nombres de voix; il en est ainsi, par exemple, dans la combinaison AABABB.....

2° Celles où B a la majorité au début; il la perd nécessairement, dans le cours du dépouillement. Cette classe comprend toutes les combinaisons qui commencent par B; par exemple, BBABAA.....

Il résulte de là que, dans toutes les combinaisons défavorables, il y a un nombre pair $2p$ des premiers tirages au bout desquels A et B sont à égalité; et l'on peut, pour une combinaison déterminée, ne considérer que le plus petit nombre $2p$ qui possède cette propriété.

Remarquons alors qu'une telle combinaison reste défavorable quand on permute A et B dans les $2p$ premiers termes; les combinaisons des deux sortes se correspondent deux à deux, et il y en a donc le même nombre qui commencent par A et par B. Tel est le résultat fondamental que l'on doit à M. Désiré André. La solution du problème est alors immédiate.

---

[1] Ce problème a été résolu pour la première fois par M. Désiré André. La solution que nous en donnons diffère légèrement de la sienne.

La probabilité pour qu'une combinaison commence par B est en effet $\frac{n}{m+n}$, la probabilité pour que A n'ait pas constamment la majorité est donc $\frac{2n}{m+n}$, ce qui donne enfin, pour la probabilité cherchée,

$$P = 1 - \frac{2n}{m+n} = \frac{m-n}{m+n}.$$

En particulier, $P = \frac{1}{2}$ si A possède les $\frac{3}{4}$ des voix.

Les problèmes de ce genre nécessitent l'emploi d'un artifice propre à chacun d'eux; quelle que soit l'ingéniosité qu'ils permettent de mettre en œuvre, ce caractere diminue leur intérêt théorique, et c'est pourquoi nous nous en tiendrons à cet exemple Nous allons étudier maintenant, sur de nouveaux problèmes, l'emploi de quelques méthodes régulières. Nous commencerons par un problème bien connu de probabilité denombrable.

5. Probleme des 3 joueurs (ou de la poule). — *On considère 3 joueurs A, B, C A et B jouent d'abord entre eux, et le gagnant de leur partie joue avec C, le gagnant de la deuxième partie joue avec le sortant de la première, et ainsi de suite Dans chaque jeu, les deux partenaires ont des chances égales, et le gagnant définitif est celui qui gagne deux parties de suite. Quelles sont les probabilités de gain de chaque joueur?*

Remarquons, tout d'abord, que la première partie se distingue des suivantes, car il est impossible qu'elle désigne un gagnant définitif Supposons alors que, dans le cours du jeu, A soit restant, B rentrant, et C sortant, et désignons par a, b, c leurs probabilités de gain.

L'application des deux principes fondamentaux, à la partie jouée entre A et B, conduit immédiatement aux relations

$$a = \frac{1}{2} + \frac{1}{2}c,$$

$$b = \frac{1}{2}a,$$

$$c = \frac{1}{2}b;$$

d'où l'on déduit les valeurs cherchées

$$a = \frac{4}{7}, \qquad b = \frac{2}{7}, \qquad c = \frac{1}{7}.$$

Il ne faut pas oublier que ces probabilités ne sont pas des rapports entre des nombres de cas favorables et un nombre de cas possibles Ce sont des sommes de progressions géométriques, comme il est facile de s'en rendre compte par un calcul direct.

La somme de ces trois probabilités est égale à l'unité; ce résultat exprime que la probabilité pour que le jeu s'éternise est nulle, ce qui n'est pas évident *a priori*, mais il ne faut pas en conclure que cette circonstance est impossible, puisque le nombre des cas possibles est infini.

Les circonstances de la première partie sont spéciales; le joueur C est assuré de rentrer pour la deuxième partie, si l'on désigne par l'indice o les probabilités relatives au début du jeu, on a donc

$$c_0 = b = \frac{2}{7}.$$

D'autre part, A et B ont la même probabilité de gain,

$$a_0 = b_0 = \frac{1}{2}(a + c) = \frac{5}{14}.$$

On voit qu'il y a un léger avantage à commencer à jouer.

Analyse du jeu. — Ces derniers résultats peuvent être obtenus directement, sous forme de progressions géométriques, par une analyse détaillée des suites possibles de parties; cette analyse va nous permettre d'ailleurs de compléter l'étude précédente.

Nous supposons toujours que la première partie se joue entre A et B; la deuxième partie sera nécessairement l'une ou l'autre des combinaisons

(2) AC, BC;

si A ou B gagne cette partie, le jeu est terminé; sinon, les joueurs font une troisième partie, qui est l'une des combinaisons

(3) CB, CA;

puis, si le jeu ne s'arrête pas, aura lieu la quatrième partie

(4) BA ou AB,

puis, au besoin, la cinquième partie

(5) AC ou BC,

et ainsi de suite. On voit immédiatement que les combinaisons possibles se reproduisent ensuite périodiquement de 3 en 3 parties.

Dans chaque combinaison, le joueur écrit à droite étant rentrant, seul le joueur de gauche peut gagner le jeu, et sa chance est $\frac{1}{2}$. Dans chaque partie à partir de la deuxième, le jeu a une chance sur deux de s'arrêter, donc la probabilité pour qu'il s'arrête à la $p^{\text{ième}}$ partie est $\frac{1}{2^{p-1}}$, il a donc la probabilité $\frac{1}{2^p}$ de se terminer sur l'une des deux combinaisons possibles de cette partie.

Ces remarques faites, cherchons les probabilités de gain a', b', c' des trois joueurs A, B, C dans les $3n$ premières parties. L'examen de la suite des parties possibles nous montre que A ne peut espérer terminer le jeu que dans les parties de rangs

$$2,\quad 4,\quad 5,\quad 7,\quad 8,\quad \ldots,\quad 3n-2,\quad 3n-1.$$

Donc sa probabilité de gagner, égale d'ailleurs à celle de B, est

$$\begin{aligned} a' = b' &= \frac{1}{2}\left[\frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^4} + \ldots + \frac{1}{2^{3n-3}} + \frac{1}{2^{3n-2}}\right] \\ &= \frac{1}{4} + \frac{3}{2^5}\left[1 + \frac{1}{2^3} + \ldots + \frac{1}{2^{3n-6}}\right] \\ &= \frac{1}{4} + \frac{3}{2^5}\,\frac{1 - \frac{1}{2^{3n-3}}}{1 - \frac{1}{2^3}} = \frac{5}{14} - \frac{3}{7\cdot 2^{3n-1}}. \end{aligned}$$

Au contraire, C ne peut terminer le jeu que dans les parties de rangs 3, 6, ..., $3n$, mais il a une chance dans les deux combinaisons possibles de ces parties, de sorte que sa probabilité de gain est

$$c' = \frac{1}{2^2} + \frac{1}{2^5} + \ldots + \frac{1}{2^{3n-1}} = \frac{1}{4}\,\frac{1 - \frac{1}{2^{3n}}}{1 - \frac{1}{2^3}} = \frac{2}{7} - \frac{1}{7\cdot 2^{3n-1}}.$$

Lorsque n augmente indéfiniment, ces probabilités, a', b', c' tendent vers les valeurs a_0, b_0, c_0 déjà trouvées autrement, mais cette

analyse du jeu par la méthode des probabilités composées est évidemment plus instructive que la méthode initiale [1].

Enfin, la probabilité pour que le jeu soit terminé au bout de $3n$ parties au plus est

$$a' + b' + c' = 1 - \frac{1}{2^{3n-1}};$$

ou encore, « la probabilité pour que le jeu ne s'arrête qu'après la $(3n)^{\text{ième}}$ partie est $\frac{1}{2^{3n-1}}$ », le calcul direct de ce résultat est d'ailleurs immédiat. Remarquons encore que, dire que la probabilité, pour que le jeu ne s'arrête jamais, est nulle, exprime que $\frac{1}{2^{3n-1}}$ tend vers zéro lorsque n augmente indéfiniment

Remarque. — Les probabilités que nous avons obtenues pour le début du jeu paraissent légèrement plus compliquées que les probabilités trouvées au cours du jeu Cependant, lorsqu'on étudie le problème plus général, relatif à $n+1$ joueurs, c'est le début qui donne les probabilités les plus simples.

$A_1, A_2, \ldots, A_{n+1}$ étant ces joueurs, on suppose que A_1 joue tout d'abord avec A_2; le gagnant joue avec A_3, et ainsi de suite, jusqu'à ce que l'un des joueurs gagne n parties successives.

Ici encore, la probabilité pour que le jeu dure indéfiniment est nulle, donc les probabilités $a_1, a_2, \ldots, a_{n+1}$ des joueurs, relatives à la première partie, vérifient la relation

$$a_1 + a_2 + \ldots + a_{n+1} = 1.$$

On trouve, d'autre part, par un calcul dont nous ne donnerons pas le détail,

$$a_1 = a_2 = \frac{1}{A}(1+2^n)^{n-1},$$

$$a_3 = \frac{1}{A}(1+2^n)^{n-2}2^n,$$

$$\ldots\ldots\ldots\ldots,$$

$$a_{p+1} = \frac{1}{A}(1+2^n)^{n-p}2^{n(p-1)},$$

$$\ldots\ldots\ldots\ldots,$$

$$a_{n+1} = \frac{1}{A}2^{n(n-1)}.$$

[1] La même méthode s'appliquerait au calcul de ces probabilités dans le cours du jeu, c'est-à-dire en supposant que A a gagné la partie précédant la première partie considérée.

Le facteur A est déterminé par la relation $a_1 + \ldots + a_{n+1} = 1$. On voit que la probabilité diminue quand le rang s'élève, mais sa décroissance est lente si n est grand ; on a, en effet,

$$\frac{a_p}{a_{p+1}} = 1 + \frac{1}{2^n},$$

et l'on voit que cette décroissance est régulière d'un joueur au suivant.

6. Espérance mathématique. — Si la production d'un événement favorable entraîne le gain d'une certaine somme, on appelle *espérance mathématique* le produit de la probabilité de l'événement par cette somme.

L'intérêt de cette notion résulte de ce que l'espérance mathématique relative à plusieurs evénements est égale à la somme des espérances relatives à chacun d'eux, qu'ils soient independants ou non.

Par exemple, si deux joueurs font trois parties successives d'un jeu égal, l'enjeu de chaque partie étant 1^{fr}, chacun d'eux a une probabilité $\frac{1}{8}$ de gagner 3^{fr} ou 0^{fr} et une probabilité $\frac{3}{8}$ de gagner 2^{fr} ou 1^{fr}

Leur espérance est donc

$$\frac{1}{8}\,3 + \frac{3}{8}(2+1) = \frac{12}{8};$$

ce résultat est bien égal à $3 . \frac{1}{2}$, somme des espérances relatives à chacune des trois parties

Nous allons appliquer cette notion à la résolution de deux problèmes bien connus.

7. Problème des rencontres — *Une urne contient n numéros, que l'on tire l'un après l'autre; il y a rencontre lorsqu'un numéro p sort au $p^{ième}$ tirage. On se propose de calculer l'espérance mathématique d'un joueur qui touche 1^{fr} à chaque rencontre; quel doit être son enjeu pour que le jeu soit équitable ?*

Tous les numéros ont la même probabilité de sortie au $p^{ième}$ tirage, donc la probabilité, et, par suite, l'espérance de la $p^{ième}$ rencontre, est $\frac{1}{n}$. L'espérance mathématique relative à l'ensemble

des n tirages est donc égale à 1^{fr}. Tel doit être l'enjeu, que le joueur perd s'il ne se produit aucune rencontre

Remarquons que le fait de gagner la première partie ne diminue pas l'espérance mathématique, alors que le fait de la perdre la diminue de $\frac{1}{n-1}$. Si le joueur gagne les $(n-1)$ premières parties, il gagne nécessairement la $n^{\text{ième}}$, donc son gain peut être n ou $n-2$, mais jamais $n-1$.

Ces raisonnements ne sont pas du tout contraires à l'affirmation que l'espérance mathématique du $p^{\text{ième}}$ tirage est $\frac{1}{n}$, en effet le joueur verse son enjeu avant le commencement du jeu; c'est $\frac{1}{n}$ franc qu'il peut vendre l'espérance d'un tirage de rang déterminé, avant le premier tirage, mais il est clair que cette valeur est modifiée dès que l'acheteur est renseigné sur des tirages antérieurs.

Ce problème peut être résolu directement en raisonnant sur les permutations de n lettres. Ces permutations peuvent être partagées en $(n-1)!$ groupes de permutations circulaires. Dans chacun de ces groupes, il y a évidemment n rencontres en tout; donc tous les tirages possibles, au nombre de $n!$, contiennent $n!$ rencontres, ce qui donne la probabilité unité d'avoir une rencontre par tirage.

Si $p_0, p_1, \ldots, p_{n-2}, p_n$ sont les probabilités de gagner $0, 1, \ldots, n$ n francs, on peut donc écrire les relations

$$p_0 + p_1 + \ldots + p_{n-2} + p_n = 1,$$
$$p_1 + 2p_2 + \ldots + (n-2)p_{n-2} + np_n = 1.$$

8. Problème de l'aiguille de Buffon (¹). — *Jetons une aiguille, au hasard, sur un plan, sur lequel sont tracées des droites parallèles équidistantes. Quelle est la probabilité pour que l'aiguille rencontre l'une quelconque de ces droites?*

Ce problème semble être du ressort de la probabilité continue. Cependant la notion d'espérance mathématique permet d'en donner une solution très simple.

Supposons que l'on touche 1^{fr} par point d'intersection. Si l'aiguille

---

(¹) Buffon est en effet le premier qui ait donné de ce problème une solution correcte, la solution ingénieuse que nous exposons ici est due à M. Barbier (*Journal du Liouville*, 1860).

est suffisamment courte, il ne peut y avoir qu'un point d'intersection au plus, et l'espérance mathématique se confond avec la probabilité.

Mais si l'aiguille est remplacée par un arc de courbe, par exemple une ligne polygonale, son espérance mathématique est la somme des espérances mathématiques de ses côtés. Si c'est une courbe plus compliquée, on peut la partager en arcs partiels, qui sont susceptibles d'être considérés séparément; en effet, l'espérance mathématique d'un élément de la courbe peut être vendue à un prix proportionnel à la longueur de cet élément.

Il suffit alors de remplacer l'aiguille par un cercle de diamètre $2a$, en désignant par $2a$ la distance des droites parallèles. La longueur de ce cercle est $2a\pi$, et son espérance mathématique, comme on s'en assure immédiatement, est rigoureusement égale à 2. L'espérance de l'unité de longueur est donc $\frac{1}{a\pi}$, et, par suite, celle d'une aiguille de longueur l est $\frac{l}{a\pi}$. Si l est inférieur à $2a$, $\frac{l}{a\pi}$ représente également la probabilité. Buffon prenait $l = a$, ce qui donne la probabilité $\frac{1}{\pi}$.

Malgré l'intérêt de la notion d'espérance mathématique, il est utile de remarquer qu'elle ne renseigne pas sur les écarts; ici, par exemple, elle ne nous apprend rien du nombre de points d'intersection que l'on peut avoir.

II. — PROBABILITÉS CONTINUES.

9. Nous savons qu'une probabilité continue peut être représentée par la position d'un point dans l'espace, ou dans un plan, ou même sur une droite.

Considérons, par exemple, une région (R) du plan, limitée par un contour simple. Étant donné un point P assujetti à être dans (R), on peut définir la probabilité pour qu'il soit dans une région (R′) intérieure à (R). Il est naturel de supposer que cette probabilité, positive par définition, varie d'une manière continue avec (R′).

Si (R′) est un élément d'aire $dx\,dy$ de la région (R), on définira la probabilité par l'expression $\varphi(x, y)\,dx\,dy$, la fonction $\varphi(x, y)$ étant positive dans (R). Bien que l'hypothèse de la continuité de la probabilité n'exige pas que $\varphi(x, y)$ soit continue, nous admettrons la continuité de cette fonction.

Enfin, d'après ce que nous avons vu dans le premier Chapitre, $\varphi(x, y)$ est assujettie à vérifier l'égalité

$$\int\int_{(R)} \varphi(x, y)\, dx\, dy = 1$$

Il résulte de là que, si on suppose infinie l'aire du domaine (R), la fonction $\varphi(x, y)$ ne peut être constante dans tout le plan; en effet, l'intégrale n'a un sens que si $\varphi(x, y)$ s'annule à l'infini

Il est possible de transformer cette probabilité continue par un changement de variables

$$x = x(X, Y), \qquad y = y(X, Y)$$

(S) et (S′) étant, dans le plan des X, Y, les domaines transformés de (R) et de (R′), la probabilité $\int\int_{R} \varphi(x, y)\, dx\, dy$ s'écrit, dans le nouveau système de variables,

$$\int\int_{(S')} \varphi(x, y) \left| \frac{D(x, y)}{D(X, Y)} \right| dX\, dY$$

Ceci revient à changer $\varphi(x, y)$ en $\varphi(x, y)\left|\frac{D(x, y)}{D(X, Y)}\right|$. On peut donc toujours effectuer un changement de variables qui, remplaçant $\varphi(x, y)$ par l'unité, permet de représenter la probabilité par l'aire (S′) elle-même.

Ce résultat peut s'obtenir par une transformation de la forme

$$X = \theta(x, y), \qquad Y = y;$$

il suffit, en effet, que l'on ait

$$\frac{\partial \theta}{\partial x} = \varphi(x, y),$$

et, par suite,

$$X = \int_0^x \varphi(x, y)\, dx, \qquad Y = y.$$

On sait que la difficulté de ces changements de variables tient au signe du déterminant fonctionnel, qui n'intervient que par sa valeur absolue. Il faut que ce déterminant ne soit ni nul ni infini dans (R); d'ailleurs, sans cela, la correspondance entre les deux aires (R) et (S) ne serait pas univoque.

Ceci n'exclut évidemment pas la possibilité de faire correspondre, à des portions du plan, des surfaces complètes; par exemple, la sphère complète à tout le plan, par projection stéréographique, ou le tore à un rectangle

Nous allons examiner quelques problèmes simples de probabilité continue, en nous bornant toujours à des problèmes du premier ordre. Auparavant, examinons la méthode générale de résolution de ces problèmes.

10. Problème général. — Considérons une aire plane (A) et un point M, de coordonnées cartésiennes (x, y), intérieur à cette aire. Nous venons de voir que l'on peut supposer que la probabilité élémentaire est proportionnelle a l'aire élémentaire $dx\,dy$, de sorte que la probabilité pour que M soit dans une aire (R), intérieure à (A), est égale au rapport $\frac{R}{A}$ des mesures de ces deux aires.

Etant donnés une deuxième aire (A′), et un point intérieur M′, de coordonnées (x', y'), on peut de même supposer que la probabilité pour que M′ se trouve dans une aire (R′) intérieure a (A′) est représentée par le rapport $\frac{R'}{A'}$.

Ceci posé, soit $\varphi(x, y; x', y')$ une fonction des deux points M et M′, que nous supposerons continue Quand M et M′ parcourent respectivement les deux aires A et A′, la fonction φ varie dans un certain intervalle (α, β). Nous nous proposons de déterminer la probabilité pour que la valeur de φ soit comprise dans un certain intervalle $(\gamma, \gamma + d\gamma)$, compris lui-même dans l'intervalle (α, β).

La probabilité cherchée est de la forme $\theta(\gamma)\,d\gamma$, et doit vérifier évidemment $\int_{\alpha}^{\beta} \theta(\gamma)\,d\gamma = 1$

Si nous fixons la position de M, les points M′, pour lesquels $\gamma \leqq \varphi \leqq \gamma + d\gamma$, sont compris entre deux arcs de courbe infiniment rapprochés, et la mesure de l'aire limitée par ces deux arcs est de la forme $F(\gamma; x, y)\,d\gamma$ Sa détermination est un problème d'analyse La probabilité pour que M′ soit intérieur à cette aire est $\frac{1}{A'}F(\gamma; x, y)\,d\gamma$.

Si M, au lieu d'être fixé en (x, y), est seulement assujetti à être intérieur à un élément d'aire $dx\,dy$ entourant (x, y), dx et dy étant infiniment petits par rapport à $d\gamma$, la modification apportée à $F(\gamma; x, y)\,d\gamma$

est négligeable; ce raisonnement est d'emploi très fréquent, et nous avons évidemment le droit de le faire, puisque $d\gamma$, seul, est donné

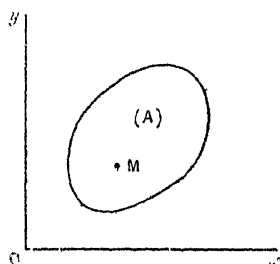
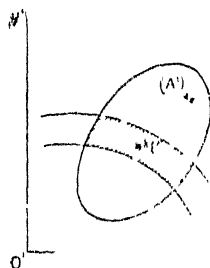



Fig. 1. Fig. 2.

Or la probabilité pour que M se trouve dans cet élément d'aire est $\frac{dx\,dy}{A}$; en vertu du principe de la probabilité composée la probabilité pour que, M étant dans cet élément d'aire, φ soit compris entre γ et $d\gamma$, est égale au produit

$$F(\gamma;\,x,\,y)\frac{d\gamma}{A'}\frac{dx\,dy}{A}.$$

Il suffit enfin d'additionner les probabilités relatives à tous les éléments d'aire de (A), pour obtenir la probabilité cherchée

$$\theta(\gamma)\,d\gamma = \frac{d\gamma}{AA'}\int\int_{(A)} F(\gamma,\,x,\,y)\,dx\,dy.$$

Nous allons employer ces généralités à résoudre quelques exemples simples relatifs à la droite et au plan.

11. Problème. — *Probabilité pour qu'un segment* MM', *intérieur à un segment de droite* AB, *de longueur* a, *ait sa longueur comprise entre* γ *et* $\gamma + d\gamma$ $(0 \leqq \gamma \leqq a)$.

Prenons comme variable la distance x du point courant M, de AB, à l'extrémité A. Si nous désignons par x' l'abscisse du point M', il faut que soit satisfaite la double inégalité

$$\gamma \leqq |x' - x| \leqq \gamma + d\gamma.$$

M étant fixé, portons de part et d'autre de ce point, sur la droite AB, $MM_1 = MM_2 = \gamma$, et considérons les éléments $d\gamma$ situés en M_1 et M_2;

M' doit se trouver dans l'un ou l'autre de ces segments élémentaires, tout en restant entre A et B. Il faut donc distinguer les diverses situations possibles de M_1 et M_2 par rapport à AB.

1° $\gamma < \frac{a}{2}$; l'un au moins des points M_1 et M_2 est intérieur à AB :
Si $0 < x < \gamma$, M_2 seul est intérieur à AB, et $F(\gamma, x) = 1$;
Si $a - \gamma < x < a$, M_1 seul est intérieur, et $F(\gamma, x) = 1$,
Si $\gamma < x < a - \gamma$, M_1 et M_2 sont tous deux intérieurs, $F(\gamma, x) = 2$.
Donc

$$\theta(\gamma) = \frac{1}{a^2}\int_0^a F(\gamma, x)\,dx = \frac{2}{a}\left(1 - \frac{\gamma}{a}\right).$$

2° $\gamma > \frac{a}{2}$; l'un au moins des points M_1 et M_2 est extérieur à AB :
Si $0 < x < a - \gamma$ ou si $\gamma < x < a$, un seul de ces points est intérieur, et $F(\gamma, x) = 1$,

A $a-\gamma$ γ B

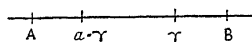



Fig. 3.

Si $a - \gamma < x < \gamma$, M_1 et M_2 sont extérieurs, et $F(\gamma, x) = 0$.
De sorte que l'on a encore

$$\theta(\gamma) = \frac{2}{a}\left(1 - \frac{\gamma}{a}\right).$$

Cette expression de $\theta(\gamma)$ est donc valable, quelle que soit la valeur de γ comprise entre 0 et a. Il est facile de vérifier que

$$\int_0^a \theta(\gamma)\,d\gamma = \frac{2}{a}\int_0^a \left(1 - \frac{\gamma}{a}\right)d\gamma$$

est égal à l'unité.

12. Le même problème peut se poser pour une aire plane. Considérons, par exemple, un carré de côté a, et proposons-nous de chercher la probabilité pour que la longueur d'un segment MM' soit comprise entre γ et $\gamma + d\gamma$.

A une position déterminée de M correspond, pour M', la couronne de centre M et de rayons γ, $\gamma + d\gamma$. Si cette couronne est intérieure, ce qui suppose, dans le cas où γ est inférieur à $\frac{a}{2}$, que M est dans le carré EFGH de côté $(a - 2\gamma)$, l'aire totale $2\pi\gamma\,d\gamma$ de cette couronne convient à M', et $F(\gamma; x, y) = 2\pi\gamma$.

Si M est dans la région comprise entre les deux carrés, la couronne ne convient pas entièrement. On voit même qu'il faut distinguer le cas où la couronne a un seul arc extérieur au carré ABCD de celui où elle en a deux; cette dernière circonstance se présente lorsque M est dans une des régions hachurées de la figure

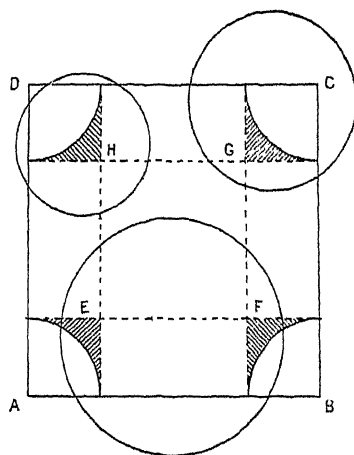



Fig 4.

Il y a donc quatre intégrales distinctes à calculer lorsqu'on

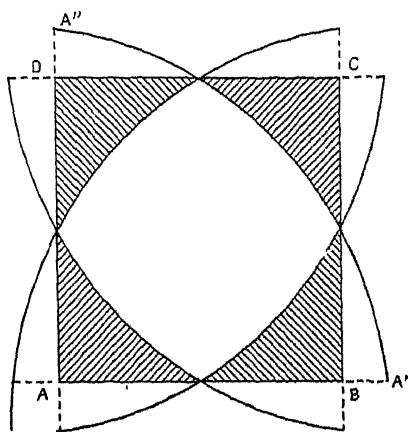



Fig. 5.

suppose $o < \gamma < \frac{a}{2}$; si l'on avait $\frac{a}{2} < \gamma < a$, ou $a < \gamma < a\sqrt{2}$, il y aurait encore à examiner les circonstances qui se présentent, suivant

la position que le point M occupe dans le carré. Le cas le plus simple est celui où $a < \gamma < a\sqrt{2}$, car il ne se présente qu'une seule espèce d'intégrales à calculer, une portion de la couronne ne convient en effet que si M est dans la région hachurée de la figure, et $F(\gamma; x, y)$ est nulle dans la région non hachurée [1] ($AA' = AA'' = \gamma$).

Nous verrons, en étudiant les problèmes du deuxième ordre, que ce genre de difficulté peut être évité pour ces problèmes, bien que ceux-ci, de prime abord, semblent plus complexes que les problèmes similaires du premier ordre.

13. Les raisonnements que l'on est amené à faire dans l'étude des probabilités continues sont parfois très délicats, et exigent certaines précautions. Un exemple célèbre de ce genre de difficulté est le « problème de Sylvester », dont nous allons dire quelques mots.

Étant donnée une région C *du plan, déterminer la probabilité pour que quatre points, pris au hasard dans cette région, déterminent un quadrilatère convexe*

Pour que quatre points M_1, M_2, M_3, M_4 ne puissent déterminer aucun quadrilatère convexe, il faut et il suffit que l'un de ces points soit intérieur au triangle formé par les trois autres. M_1, M_2, M_3 étant donnés, la probabilité pour que M_4 soit intérieur à leur triangle est égale au rapport $\frac{T_4}{A}$ des aires de ce triangle et de C

Si l'on se donne les quatre points, et si T_1, T_2, T_3, T_4 sont les aires des quatre triangles qu'ils déterminent trois à trois, la probabilité pour qu'ils ne définissent pas de quadrilatère convexe est $\frac{T_1 + T_2 + T_3 + T_4}{A}$. Si T est la moyenne des aires des quatre triangles, cette probabilité se mesure encore par le rapport $\frac{4T}{A}$.

Soient alors $d\omega_1$, $d\omega_2$, $d\omega_3$, $d\omega_4$ les aires infinitésimales associées respectivement à M_1, M_2, M_3, M_4, la probabilité cherchée est donnée par l'intégrale multiple

$$\int_c \left(1 - \frac{4T}{A}\right) d\omega_1\, d\omega_2\, d\omega_3\, d\omega_4.$$

---

[1] Sur la figure 5, on a pris $\gamma = a\sqrt{2}$.

On conçoit que le calcul soit très pénible, même s'il s'agit de contours élémentaires. C'est pour le cercle que cette probabilité est maximum, ce maximum est un nombre très voisin de 0,704 [1].

Dès que l'on cherche à se poser le même problème pour tout le plan, on se heurte à des difficultés dont nous allons dire quelques mots. Donnons-nous arbitrairement les trois points M_1, M_2, M_3, pour que le quatrième point M_4 détermine, avec ces trois points, un quadrilatère convexe, il faut qu'il se trouve dans une des régions hachurées de la figure.

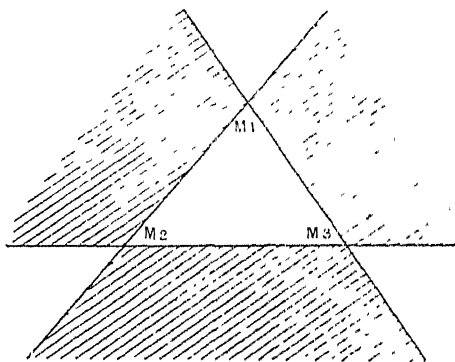



Fig. 6.

On peut remplacer le contour C du problème précédent par un cercle de rayon infiniment grand, ce qui conduit à adopter, pour mesure de la probabilité cherchée, le nombre $\frac{1}{2}$. Mais ce raisonnement suppose que M_1, M_2, M_3 sont à distance finie, et que, d'autre part, M_4 a la même probabilité de se trouver dans le voisinage de ces trois points que d'en être infiniment loin, ces deux hypothèses, faites simultanément, sont évidemment contradictoires, et font comprendre, sans plus amples explications, que le problème ainsi posé n'a pas de sens précis.

III. — PROBABILITÉS DÉNOMBRABLES [2].

14. On peut considérer trois catégories de probabilités dénom-

---

(1) On trouvera de plus amples détails sur ce problème, et en particulier, sur l'étude approfondie du cercle, dans la thèse de M. Deltheil.

(2) *Cf.* Borel, *Sur les probabilités dénombrables* (*Rendiconti del Circolo Math. di Palermo*, t. XXVII).

brables. Dans les deux premières. l'épreuve comprend une infinité dénombrable de cas possibles; les épreuves peuvent être en nombre fini, comme dans les problèmes de la première catégorie, ou en nombre infini, comme dans ceux de la deuxième catégorie. Les probabilités des cas possibles d'une même épreuve forment évidemment une série convergente dont la somme est egale à l'unite

Nous ne nous occuperons ici que des problèmes de la troisième catégorie; ce sont les problèmes qui reposent sur une infinité dénombrable d'épreuves, ne comprenant chacune qu'un nombre fini de cas possibles

Le cas le plus simple est celui où le nombre des cas possibles est egal à 2 Désignons par $p_1, p_2, \ldots, p_n, \ldots$ les probabilités, toutes comprises entre o et 1, des événements favorables dans les épreuves de rang 1, 2, ..., n, ... Les événements contraires ont respectivement les probabilités

$$q_1 = 1 - p_1, \quad q_2 = 1 - p_2, \quad \ldots, \quad q_n = 1 - p_n, \quad \ldots$$

Nous écartons la circonstance où l'une des épreuves comporterait une certitude ($p_n = 0$ ou 1), car il est clair que l'on peut supprimer une pareille épreuve sans modifier la nature du problème.

Dans ces conditions, supposons réalisée une suite infinie d'epreuves, et proposons-nous de déterminer la probabilité pour que le cas favorable se produise un nombre donné de fois.

Désignons par $A_0, A_1, \ldots, A_n, \ldots$ les probabilités d'obtenir l'événement favorable, o, 1, ..., n, .. fois. Il faut, d'autre part, introduire la probabilité de réaliser une infinité de cas favorables; désignons-la par A_∞. Il n'est pas inutile d'en préciser la signification Remarquons tout d'abord que la série $A_0 + A_1 + \ldots + A_n + \ldots$ est convergente, et inférieure ou égale à 1. Soit $P_m^{(n)} (m \leqq n)$ la probabilité de réaliser m événements favorables dans les n premières épreuves On a, évidemment,

$$P_0^{(n)} + P_1^{(n)} + \ldots + P_n^{(n)} = 1.$$

A_m est, par définition, la limite de $P_m^{(n)}$ lorsque n augmente indéfiniment. Or, la probabilité d'obtenir plus de m cas favorables dans les n premières épreuves ($n > m$) est

$$P_{m+1}^{(n)} + P_{m+2}^{(n)} + \ldots + P_n^{(n)} = 1 - [P_0^{(n)} + P_1^{(n)} + \ldots + P_m^{(n)}];$$

lorsque n augmente indéfiniment, cette probabilité devient donc

$$1-(A_0+A_1+\ldots+A_m).$$

Si l'on suppose maintenant que m augmente indéfiniment, cette dernière expression, positive et non croissante, a nécessairement une limite

$$A_\infty = 1-(A_0+A_1+A_2+\ldots),$$

que l'on peut considérer, avec un sens maintenant bien précis, comme la probabilité d'obtenir une infinité de cas favorables.

Nous allons déterminer les probabilités successives $A_0, A_1, A_2, \ldots$, ce qui nous montrera que le résultat dépend essentiellement de la nature de la série $p_1+p_2+\ldots+p_n+\ldots$

1° *Calcul de* A_0. — On a, évidemment,

$$P_0^{(n)} = q_1 q_2 \ldots q_n = (1-p_1)(1-p_2)\ldots(1-p_n).$$

Si la série Σp_n est convergente, il en est de même du produit infini $\Pi(1-p_n)$; sa limite, bien déterminée et non nulle, est la valeur cherchée de A_0.

Si la série Σp_n est divergente, le produit infini est divergent, et A_0 est nul. Ce résultat ne signifie nullement l'impossibilité de n'obtenir aucun cas favorable, mais seulement le fait que la probabilité de n'en obtenir aucun est d'autant plus voisine de zéro que le nombre des épreuves est plus grand.

2° *Calcul de* A_1. — La probabilité pour que l'unique événement favorable obtenu soit le $m^{\text{ième}}$ est mesurée par

$$\lim_{n=\infty} q_1 \ldots q_{m-1} p_m q_{m+1} \ldots q_n = A_0 \frac{p_m}{1-p_m},$$

en admettant que la série (p_n) est convergente.

L'addition des probabilités relatives à tous les rangs possibles de l'événement favorable donne

$$A_1 = A_0\left(\frac{p_1}{1-p_1}+\frac{p_2}{1-p_2}+\ldots\right).$$

La série entre parenthèses est convergente avec Σp_n, de sorte que A_1 a une valeur bien déterminée.

Si la série Σp_n est divergente, on peut considérer un nombre fini n

d'épreuves, pour lequel on a

$$P_1^{(n)} = (1-p_1)(1-p_2)\ldots(1-p_n)\left[\frac{p_1}{1-p_1}+\frac{p_2}{1-p_2}+\ldots+\frac{p_n}{1-p_n}\right].$$

Cette expression se présente sous une forme indéterminée, quand n augmente indéfiniment, car la parenthèse $\sum_{i=1}^{n}\frac{p_i}{1-p_i}$, infiniment grand de l'ordre de $\sum_{i=1}^{n} p_i$, est multipliée par un facteur infiniment petit [1]. Au lieu de l'égalité, considérons l'inégalité qu'elle entraîne

$$P_1^{(n)} < e^{-(p_1+p_2+\ldots+p_n)}\left(\frac{p_1}{1-p_1}+\frac{p_2}{1-p_2}+\ldots+\frac{p_n}{1-p_n}\right),$$

le deuxième membre, de l'ordre de $e^{-S_n}S_n$ tend vers zéro quand S_n augmente indéfiniment avec n. Donc A_1 est nul si la série Σp_n est divergente. Ce résultat paraît naturel, si l'on remarque que la probabilité, de gagner une partie de rang déterminé m, n'est autre que $p_n A'_0$, A'_0 étant la valeur que prend A_0 lorsque, dans la suite d'épreuves, on supprime la $m^{\text{ième}}$ épreuve, cet A_0 est nul, car la série des p_n restants est encore divergente.

3° *Calcul de* A_2. — A_2 se calcule par le même raisonnement que A_1. On trouve

$$A_2 = A_0 \sum_{i,k=1}^{\infty} \frac{p_i p_k}{(1-p_i)(1-p_k)} \qquad (i \neq k)$$

A_2 a une valeur finie non nulle si (p_n) est convergente; elle est nulle, au contraire, si (p_n) est divergente.

En résumé, si la série des p_n est divergente, tous les A_m sont nuls, de sorte que la probabilité A_∞ est égale à 1.

Si la série (p_n) est convergente, on a, d'une manière générale,

$$A_p = A_0 \sum_{i_1 i_2 \ldots i_p} \frac{p_{i_1} p_{i_2}\ldots p_{i_p}}{(1-p_{i_1})(1-p_{i_2})\ldots(1-p_{i_p})},$$

---

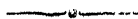
[1] $\sum \frac{p_n}{1-p_n}$ et Σp_n ne sont du même ordre que si p_n ne tend pas vers 1, quand n augmente indéfiniment, mais le résultat $A_1 = 0$ est vrai *a fortiori* dans ce cas

cette somme étant étendue à toutes les combinaisons de p indices distincts.

On a alors immédiatement

$$A_0 + A_1 + \ldots + A_p + \ldots = A_0\left[1 + \sum \frac{p_i}{q_i} + \sum \frac{p_i p_k}{q_i q_k} + \ldots\right]$$
$$= A_0\left(1 + \frac{p_1}{q_1}\right)\left(1 + \frac{p_2}{q_2}\right)\cdots\left(1 + \frac{p_n}{q_n}\right)\cdots$$
$$= \frac{A_0}{q_1 q_2 \ldots q_n} = 1.$$

Dans ce cas, A_∞ est nul.

Il résulte de là qu'il est impossible de construire une suite dénombrable de probabilités pour laquelle la probabilité d'avoir une infinite de cas favorables serait différente de 0 ou 1

CHAPITRE III.

PROBABILITÉS DISCONTINUES. PROBLÈMES DU DEUXIEME ORDRE.

— LOI DE GAUSS

1. Nous allons reprendre le probleme général des probabilités discontinues qui va nous conduire à la loi remarquable de Gauss.

Nous avons vu que, si la probabilité du cas favorable P d'une certaine épreuve est p, $q = 1 - p$ étant, par suite, la probabilité du cas défavorable Q, la probabilité d'obtenir α cas favorables et $n - \alpha$ cas defavorables dans une suite de n épreuves, est $\frac{n!}{\alpha!(n-\alpha)!} p^\alpha q^{n-\alpha}$.

Nous nous proposons ici de calculer une valeur approchée de cette expression, lorsque le nombre n des épreuves est tres grand, ainsi que α et $n - \alpha$ Cette question se rattache à l'etude de la croissance de la fontion e^n. On sait, en effet, que la fonction

$$e^n = 1 + \frac{n}{1} + \frac{n^2}{2!} + \ldots + \frac{n^{n-1}}{(n-1)!} + \frac{n^n}{n!} + \ldots$$

augmente indéfiniment avec n, et que sa croissance est plus rapide que celle de toute puissance de n

Lorsque n augmente indéfiniment, un nombre de plus en plus grand de termes de cette série deviennent très grands, mais le point fondamental est que la rapidité de croissance de e^n est due uniquement a l'influence d'un petit nombre de termes de cette série. C'est ainsi que les deux termes les plus grands, $\frac{n^{n-1}}{(n-1)!}$ et $\frac{n^n}{n!}$, d'ailleurs égaux entre eux, fournissent déjà une valeur relativement bonne pour e^n, si n est très grand.

Tout d'abord, on a l'inégalité évidente

$$e^n > \frac{n^n}{n!},$$

ou encore

$$(1) \qquad n! > \left(\frac{n}{e}\right)^n.$$

2. Déterminons, d'autre part, une approximation par excès de $n!$ Pour cela, mettons en évidence, dans le développement de e^n, le terme le plus élevé $\frac{n^n}{n!}$. Nous avons

$$e^n = \frac{n^n}{n!}\left[1 + \frac{n}{n+1} + \frac{n^2}{(n+1)(n+2)} + \ldots + \frac{n^p}{(n+1)(n+2)\ldots(n+p)} + \ldots \right.$$
$$\left. + 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \ldots + \frac{(n-1)!}{n^{n-1}}\right],$$

la première ligne du second membre comprenant les termes de la série qui suivent $\frac{n^n}{n!}$, et, la deuxième ligne, les termes qui le précèdent.

Si n est très grand, les premiers termes de ces deux lignes diffèrent peu de 1, en particulier, il en est ainsi des n' premiers termes, en désignant par n' la partie entière de $\sqrt{n}$.

Examinons, en effet, la décroissance du terme général de la première ligne.

$$u_p = \frac{n^p}{(n+1)(n+2)\ldots(n+p)} = \frac{1}{\left(1+\frac{1}{n}\right)\left(1+\frac{2}{n}\right)\cdots\left(1+\frac{p}{n}\right)}$$

u_p est inférieur à $\frac{1}{1+\frac{p(p+1)}{2n}}$; si p est de l'ordre de $\sqrt{n}$, cette limite supérieure est de l'ordre de $\frac{2}{3}$; donc si $p > \sqrt{n}$, on a $u_p < \frac{2}{3}$; si $p > 2\sqrt{n}$, $u_p < \frac{1}{3}$, et l'on voit, d'une manière générale, que u_p décroît très rapidement quand $\frac{p}{\sqrt{n}}$ croît.

D'une manière plus précise, on a

$$\frac{u_{p+n'}}{u_p} = \frac{1}{\left(1+\frac{p+1}{n}\right)\left(1+\frac{p+2}{n}\right)\cdots\left(1+\frac{p+n'}{n}\right)}$$
$$< \frac{1}{1+\frac{(p+1)+(p+2)+\ldots+(p+n')}{n}}$$
$$< \frac{1}{\frac{3}{2}+p} < \frac{2}{3};$$

il résulte de là que la somme des termes de la première ligne est inférieure à la somme des n' premiers termes, multipliée par la somme de la progression géométrique

$$1 + \frac{2}{3} + \left(\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^3 + \ldots = 3,$$

elle est donc, *a fortiori*, inférieure à $3\sqrt{n}$.

Le terme général de la deuxième ligne est

$$v_p = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{p}{n}\right) \qquad p < n),$$

et l'on a

$$\frac{v_{p+n'}}{v_p} = \left(1 - \frac{p+1}{n}\right)\left(1 - \frac{p+2}{n}\right)\cdots\left(1 - \frac{p+n'}{n}\right)$$

$$< e^{-\left[p + \frac{n'(n'+1)}{2n}\right]} < e^{-\left(p+\frac{1}{2}\right)} < \frac{1}{e}.$$

Cette ligne a donc une somme inférieure au produit, par $\frac{1}{e}$, de la somme des n' premiers termes, donc inférieure, *a fortiori*, à $\frac{\sqrt{n}}{e}$.

En résumé, il existe un nombre fini k tel que l'on ait l'inégalité

$$(2) \qquad e^n < k\sqrt{n}\,\frac{n^n}{n!}.$$

3. Les inégalités (1) et (2) se résument dans la double inégalité

$$\left(\frac{n}{e}\right)^n < n! < k\sqrt{n}\left(\frac{n}{e}\right)^n.$$

$\sqrt{n}$ croît très lentement par rapport à $\left(\frac{n}{e}\right)^n$, ce qui montre bien que cette dernière expression est une bonne valeur approchée de $n!$ lorsque n est très grand.

Si nous posons

$$(3) \qquad n! = \left(\frac{n}{e}\right)^n \sqrt{n}\,\varphi(n),$$

nous pouvons affirmer que la fonction $\varphi(n)$ reste finie quand n augmente indéfiniment; on vérifie alors immédiatement que $\frac{\varphi(n+1)}{\varphi(n)}$ tend vers 1 comme $e^{-\frac{1}{12n^2}}$, et, par suite, que $\varphi(n)$ tend vers une

limite, on a, en effet,

$$\log \varphi(n+1) - \log \varphi(n) = 1 - \left(n + \frac{1}{2}\right) \log\left(1 + \frac{1}{n}\right) = -\frac{1}{12 n^2} + \ldots$$

Nous allons déduire la détermination de cette limite de l'étude de l'intégrale

$$I_m = \int_0^{\frac{\pi}{2}} \cos^m x \, dx,$$

où m est un nombre entier positif. Le calcul de cette intégrale résulte d'une formule de récurrence, fournie par l'intégration par parties ; cette méthode donne, en effet,

$$I_m = [\sin x \cos^{m-1} x]_0^{\frac{\pi}{2}} + (m-1) \int_0^{\frac{\pi}{2}} \cos^{m-2} x \sin^2 x \, dx,$$

donc, si $m > 1$,

$$\begin{aligned} I_m &= (m-1) \int_0^{\frac{\pi}{2}} \cos^{m-2} x (1 - \cos^2 x) \, dx \\ &= (m-1)(I_{m-2} - I_m), \end{aligned}$$

d'où, en définitive,

$$m I_m = (m-1) I_{m-2}.$$

Cette formule de récurrence détermine immédiatement I_m en fonction de I_0 ou I_1, suivant que m est pair ou impair. Or

$$I_0 = \int_0^{\frac{\pi}{2}} dx = \frac{\pi}{2},$$

$$I_1 = 1,$$

d'où l'on déduit enfin

$$I_{2q} = \frac{1.3.5 \ldots 2q-1}{2.4.6 \ldots 2q} \frac{\pi}{2}$$

et

$$I_{2q+1} = \frac{2.4.6 \ldots 2q}{3.5.7 \ldots (2q+1)}.$$

On a les inégalités évidentes

$$I_{2q} > I_{2q+1} > I_{2q+2},$$

ou encore

$$1 > \frac{I_{2q+1}}{I_{2q}} > \frac{I_{2q+2}}{I_{2q}}.$$

D'autre part, le rapport

$$\frac{I_m}{I_{m-2}} = \frac{m-1}{m}$$

tend vers 1 quand m augmente indéfiniment, il résulte de là, ainsi que de la double inégalité précédente, que $\frac{I_{2q+1}}{I_{2q}}$ tend vers 1 quand m augmente indéfiniment.

Or, ce rapport contient $\frac{\pi}{2}$ et des factorielles, formons-le, il vient

$$\lim_{q=\infty} \frac{[2.4.6\ldots 2q]^2}{[1.3.5\ldots(2q-1)]^2(2q+1)} \frac{2}{\pi} = 1$$

ou

$$\frac{\pi}{2} = \lim_{q=\infty} \frac{[2.4.6\ldots 2q]^4}{[(2q)!]^2(2q+1)} = \lim_{q=\infty} \frac{2^{4q}(q!)^4}{(2q+1)[(2q)!]^2}.$$

C'est la formule bien connue de Wallis. Il suffit de remplacer maintenant les factorielles par leur expression (3) pour obtenir la limite A de $\varphi(n)$. On obtient ainsi

$$\frac{\pi}{2} = \lim_{q=\infty} \frac{2^{4q}\left(\frac{q}{e}\right)^{4q} q^2 \varphi(q)^4}{(2q+1)\left(\frac{2q}{e}\right)^{4q} 2q\,\varphi(2q)^2} = \frac{A^2}{4},$$

d'où

$$A = \sqrt{2\pi}.$$

En définitive, on peut écrire

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\,(1+\varepsilon_n), \tag{4}$$

ε_n tendant vers 0 avec $\frac{1}{n}$. Cette formule fondamentale est due à Stirling. Une plus grande approximation montre que ε_n est équivalent à $\frac{1}{12n}$.

Par exemple, pour $n = 100$, ε_n est inférieur à $\frac{1}{1000}$, et $\left(\frac{n}{e}\right)^n \sqrt{2\pi n}$ est approché avec une erreur relative inférieure à $\frac{1}{1000}$.

4. Problème général des probabilités discontinues. — Le problème que nous avons repris au début de ce Chapitre se généralise immédia-

tement, lorsque chaque épreuve contient plus de deux événements. Pour simplifier l'exposition, considérons des épreuves à trois événements, tout en conservant, à la forme du raisonnement et au calcul, toute leur généralité.

Soient P, Q, R ces événements, de probabilités p, q, r, avec $p+q+r=1$. Dans une suite de n expériences, la probabilité d'obtenir α, β, γ fois $(\alpha+\beta+\gamma=n)$, les événements P, Q, R, est donnée par le terme général

$$\text{(5)} \qquad P = p^\alpha q^\beta r^\gamma \frac{n!}{\alpha!\beta!\gamma!}$$

du développement de $(p+q+r)^n$

Nous supposerons que α, β, γ sont suffisamment grands pour que les approximations que nous allons faire soient légitimes, et nous allons chercher une valeur approximative de P.

On constate immédiatement que le terme le plus grand du développement de $(p+q+r)^n$ correspond aux valeurs de α, β, γ qui sont le plus voisines de np, nq, nr. On est ainsi conduit à poser

$$\alpha = np + x,$$
$$\beta = nq + y,$$
$$\gamma = nr + z,$$

x, y, z s'appellent les « écarts », et leur somme $x+y+z$ est nulle

Ceci établi, la formule de Stirling nous permet d'écrire

$$\log n! = n \log n - n + \frac{1}{2}\log(2\pi n) + \xi_n,$$

ξ_n étant un infiniment petit équivalent à $\frac{1}{12n}$.

Supposons que ξ_n soit négligeable dans cette formule, ainsi que les termes similaires, dans les expressions de $\log\alpha!$, $\log\beta!$, $\log\gamma!$. On peut alors écrire, en prenant les logarithmes des deux membres de (5), et en faisant les approximations admises,

$$\begin{aligned}\log P = {} & \alpha\log p + \beta\log q + \gamma\log r \\ & + n\log n - \alpha\log\alpha - \beta\log\beta - \gamma\log\gamma \\ & + \frac{1}{2}\log 2\pi n - \frac{1}{2}\log 2\pi\alpha - \frac{1}{2}\log 2\pi\beta - \frac{1}{2}\log 2\pi\gamma \\ & - n + \alpha + \beta + \gamma\end{aligned}$$

Les termes de la dernière ligne ont une somme nulle; la deuxième

ligne s'écrit encore

$$(\alpha+\beta+\gamma)\log n-\alpha\log\alpha-\beta\log\beta-\gamma\log\gamma$$
$$=\alpha\log\frac{n}{\alpha}+\beta\log\frac{n}{\beta}+\gamma\log\frac{n}{\gamma},$$

ce qui donne

$$\log P=-\alpha\log\frac{\alpha}{np}-\beta\log\frac{\beta}{nq}-\gamma\log\frac{\gamma}{nr}+\frac{3}{2}\log n-\frac{1}{2}\log\alpha-\frac{1}{2}\log\beta-\frac{1}{2}\log\gamma$$
$$+\frac{1}{2}(1-1-1-1)\log 2\pi n,$$

ou, en désignant par σ le nombre d'événements de chaque épreuve, nombre égal ici à 3,

$$\text{(6)}\quad \log P=-\left(\alpha+\frac{1}{2}\right)\log\frac{\alpha}{np}-\left(\beta+\frac{1}{2}\right)\log\frac{\beta}{nq}-\left(\gamma+\frac{1}{2}\right)\log\frac{\gamma}{nr}$$
$$-\frac{1}{2}\log pqr+\frac{1-\sigma}{2}\log 2\pi n.$$

Sous cette forme, cette formule possède tout son caractère de généralité, quel que soit le nombre σ.

Posons

$$\text{(7)}\quad \log P'=-\left(\alpha+\frac{1}{2}\right)\log\frac{\alpha}{np}-\left(\beta+\frac{1}{2}\right)\log\frac{\beta}{nq}-\left(\gamma+\frac{1}{2}\right)\log\frac{\gamma}{nr};$$

les termes du second membre sont les seuls termes de $\log P$ qui dépendent de α, β, γ, et l'on a

$$\text{(8)}\qquad P=\frac{1}{\sqrt{pqr}}\,\frac{1}{(\sqrt{2\pi n})^{\sigma-1}}\,P'.$$

Évaluons maintenant P' en introduisant les écarts. Nous avons

$$\log\frac{\alpha}{np}=\log\left(1+\frac{x}{np}\right).$$

Nous supposons que p, q, r ne sont pas très petits, de sorte que les valeurs de x, y, z, très petites par rapport à n, qui correspondent aux probabilités maxima, rendent très petits les termes $\frac{x}{np}$, $\frac{y}{nq}$, $\frac{z}{nr}$.

En négligeant les puissances quatrièmes de ces quantités, on peut donc écrire

$$\log\frac{\alpha}{np}=\frac{x}{np}-\frac{x^2}{2n^2p^2}+\frac{x^3}{3n^3p^3},$$

d'où

$$\left(\alpha+\frac{1}{2}\right)\log\frac{\alpha}{np} = \left(np+x+\frac{1}{2}\right)\left(\frac{x}{np}-\frac{x^2}{2n^2p^2}+\frac{x^3}{3n^3p^3}\right)$$
$$= x + \frac{x}{2np} + \frac{x^2}{2np} - \frac{x^2}{4n^2p^2} - \frac{x^3}{6n^2p^2},$$

les termes négligés sont de l'ordre de $\left(\frac{x}{np}\right)^3$.

$\log P'$ s'obtient en additionnant les expressions semblables, relatives à x, y, z, en se rappelant que $x+y+z=0$, ce qui donne, sous une forme qui conviendrait d'ailleurs au cas général,

$$(9)\quad \log P' = -\frac{1}{2n}\left(\frac{x^2}{p}+\frac{y^2}{q}+\frac{z^2}{r}+\frac{x}{p}+\frac{y}{q}+\frac{z}{r}\right)+\frac{1}{6n^2}\left(\frac{x^3}{p^2}+\frac{y^3}{q^2}+\frac{z^3}{r^2}\right)$$

C'est cette formule ([1]) qui, avec (8), répond à la question que nous nous étions proposée.

5. Nous allons appliquer ces résultats généraux au cas restreint où l'épreuve comporte deux événements, ce qui nous fournira la valeur approximative cherchée de P.

Ici, en effet, nous avons $y=-x$, $q=1-p$, et

$$\alpha = np + x,$$
$$\beta = nq - x,$$

d'où l'on déduit

$$(10)\qquad P = \frac{1}{\sqrt{2\pi npq}}\, e^{-\frac{x^2}{2npq}+\frac{(q-p)x^3}{6n^2p^2q^2}-\frac{(q-p)x}{2npq}}$$

C'est la probabilité pour que l'écart ait une valeur déterminée x. On voit qu'elle n'est pas rigoureusement la même pour x et $-x$, ceci tient à ce que la formule du binome n'est pas symétrique si p est différent de q. Cependant P est presque symétrique si le terme impair $\frac{(q-p)x^3}{6n^2p^2q^2}-\frac{(q-p)x}{2npq}$ est négligeable, ce qui a lieu si x est de l'ordre de $\sqrt{n}$, c'est-à-dire au voisinage du maximum de la probabilité.

([1]) Dans la formule (9) sont négligés les termes en $\frac{x^2}{n^2p^2}$, négligeables en effet par rapport aux termes en $\frac{x^3}{n^2p^2}$, si l'écart x est grand, et négligeables en même temps que ces termes dans le cas contraire.

Posons alors

$$x = t\sqrt{2npq}$$

$u = \sqrt{2npq}$ s'appelle « l'unité d'écart », et t, « l'écart relatif ». Il est bon de remarquer que u peut ne pas convenir lui-même comme écart, car $np + u$ n'est pas nécessairement un nombre entier .

En introduisant la variable t à la place de x, nous avons

$$\text{(10')} \qquad \mathrm{P} = \frac{1}{\sqrt{\pi}\,u}\, e^{-t^2 + \frac{q-p}{u}\left(\frac{2}{3}t^3 - t\right)}.$$

Si n est très grand, t restant fini, cette formule est sensiblement symétrique en t. D'ailleurs, l'erreur relative commise en négligeant le terme impair ne devient appréciable que si t est assez grand, mais alors e^{-t^2} est très petit, et l'erreur absolue est negligeable.

Nous supposerons n très grand, ce qui nous permettra de ne conserver que le terme symétrique Remarquons qu'à deux valeurs entières de x, assez voisines, correspondent deux valeurs de t très peu différentes Donnons alors à t un accroissement très petit Δt, de la valeur t' à la valeur t'' Nous avons

$$x' = t'\sqrt{2npq},$$
$$x'' = t''\sqrt{2npq},$$

et

$$\alpha'' - \alpha' = x'' - x' = (t'' - t')\sqrt{2npq} = u\Delta t$$

doit être un nombre entier.

La probabilité pour que x soit compris entre x' et x'' est

$$\mathrm{P}(t)(x'' - x') = \mathrm{P}\,u\,\Delta t,$$

$\mathrm{P}(t)$ étant la valeur de P lorsque t reste dans l'intervalle (t', t''). Il résulte de là que la probabilité pour que t soit compris entre t et $t + dt$ est

$$\frac{1}{\sqrt{\pi}\,u}\, e^{-t^2}\, u\, dt,$$

ou

$$\text{(11)} \qquad \frac{1}{\sqrt{\pi}}\, e^{-t^2} dt$$

C'est la loi fondamentale de Gauss; fait remarquable, ce résultat, pour lequel on a supposé seulement que n est très grand, est indépendant de la valeur de n.

La probabilité pour que t soit compris entre deux valeurs distinctes t' et t'' est donnée par l'intégrale

$$\frac{1}{\sqrt{\pi}}\int_{t'}^{t''} e^{-t^2}\,dt,$$

et x est alors compris entre les valeurs $x' = ut'$ et $x'' = ut''$.

Une autre propriété remarquable de cette expression de la probabilité est qu'elle convient à l'expression d'une probabilité continue, car l'intégrale précédente, prise entre $-\infty$ et $+\infty$, a pour valeur l'unité [1].

Ce résultat ne pouvait pas être prévu, étant données les approximations que nous avons faites. La formule $\int_{t'}^{t''}\frac{e^{-t^2}}{\sqrt{\pi}}\,dt$ est d'ailleurs tout à fait inexacte pour certaines valeurs de t, puisqu'elle n'a une signification que pour

$$0 \leq \alpha \leq n,$$

c'est-à-dire

$$-np \leq x \leq nq,$$

ou

$$-\sqrt{\frac{np}{2q}} \leq t \leq \sqrt{\frac{nq}{2p}};$$

il est clair qu'en dehors de cet intervalle la probabilité est rigoureusement nulle, alors que l'intégrale donne une valeur positive, donc trop grande. C'est d'ailleurs parce que, à l'intérieur de cet intervalle, elle fournit des valeurs trop petites, que la compensation peut se produire entre les limites infinies.

6. t peut être positif ou négatif; mais la probabilité dont on a besoin en général est celle pour que sa valeur absolue soit comprise entre deux nombres positifs donnés λ', λ''. D'après la symétrie de la formule adoptée, cette probabilité est représentée par

$$P_{\lambda',\lambda''} = \frac{2}{\sqrt{\pi}}\int_{\lambda'}^{\lambda''} e^{-\lambda^2}\,d\lambda = \frac{2}{\sqrt{\pi}}\left\{\int_0^{\lambda''} e^{-\lambda^2}\,d\lambda - \int_0^{\lambda'} e^{-\lambda^2}\,d\lambda\right\}.$$

On introduit la fonction

$$\Theta(\lambda) = \frac{2}{\sqrt{\pi}}\int_0^{\lambda} e^{-\lambda^2}\,d\lambda,$$

[1] Ce résultat bien connu est désigné sous le nom de *formule de Poisson*. (*Cf.* par exemple D'ADHÉMAR, *Leçons sur les principes de l'Analyse*, t. I, Chap. IV.)

dont les valeurs sont fournies par des tables, ce qui permet d'écrire

$$P_{l',l''} = \Theta(\lambda'') - \Theta(\lambda')$$

C'est cette formule que l'on emploie dans la pratique. $\Theta(\lambda)$ tend très rapidement vers sa limite 1, lorsque λ augmente. Déjà pour $\lambda = 4$ ou 5, on a $\Theta(\lambda) = 0,999999\ldots$. Donc la probabilité pour que t soit compris entre deux nombres supérieurs à 5 est très faible; il résulte de cette remarque que (11) fournit une bonne approximation de la valeur plus rigoureuse représentée par (10'), tant que u n'est pas inférieur à 5^3, c'est-à-dire à 125.

Remarque. — Il ne faut pas oublier que le terme $\frac{x}{2npq}$, dans la formule (10), n'est plus négligeable par rapport à $\frac{x^2}{2npq}$ si x est petit par rapport à $\sqrt{n}$. Si l'on ne néglige que le terme $\frac{x^3}{2n^2p^2q^2}$, P est symétrique par rapport à la valeur $x = -\frac{q-p}{2}$, et non plus par rapport à $x = 0$.

Prenons, comme exemple, une partie de 633 coups de dés avec $p = \frac{1}{6}$, $q = \frac{5}{6}$, alors $np = \frac{633}{6} = 105,5$, mais on n'a pas tout à fait la même probabilité d'amener, 105 fois et 106 fois, un numéro donné. En effet, $\frac{q-p}{2}$ est égal à $\frac{1}{3} = 0,333$, et l'axe de symétrie de la courbe $P = P(x)$

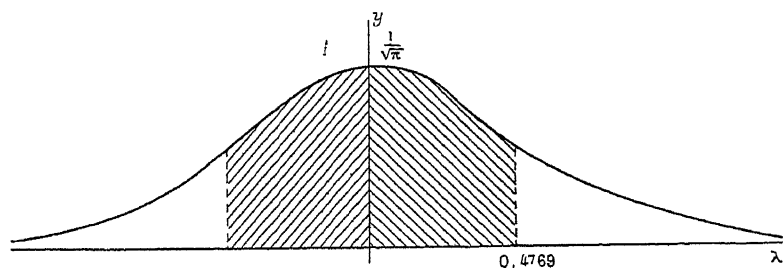



Fig. 7.

correspond à $\alpha = 105,5 - 0,33 = 105,17$ de sorte que la probabilité est légèrement plus grande de gagner 105 épreuves que d'en gagner 106.

7. Étude de la fonction $\Theta(\lambda)$. — Cette fonction est représentée par

l'aire comprise entre l'axe des λ et la courbe $y = \frac{1}{\sqrt{\pi}} e^{-\lambda^2}$. La probabilité pour que λ soit inférieur à λ' en valeur absolue est mesurée par l'aire comprise entre les deux abscisses + λ' et − λ'.

En particulier, $\Theta(\lambda') = \frac{1}{2}$ si $\lambda' = 0,4769$..., l'aire correspondante, hachurée sur la figure, est la moitié de l'aire totale. Il y a donc une chance sur deux pour que l'écart x soit inférieure à $0,4769\sqrt{2npq}$ en valeur absolue. Cet écart $0,4769u$ s'appelle « l'écart médian », et 0,4769..., « l'écart relatif médian ».

Par exemple, pour 600 parties de dés, np est égal à 100, et l'écart médian est égal à $0,4769\sqrt{\frac{500}{3}}$, ou 6 environ. On a donc une chance sur deux d'amener un numéro donné, un nombre de fois compris entre 106 et 94, on a une chance sur quatre de l'amener plus de 106 fois, et également une chance sur quatre de l'amener moins de 94 fois.

Un examen attentif des valeurs numériques de Θ(λ) conduit à une remarque curieuse. Reproduisons, en effet, les quelques valeurs suivantes, tirées de la Table du Traité de Joseph Bertrand, en arrondissant les valeurs de λ.

Θ(λ)	λ	Valeurs arrondies de λ
0,9	1,163	1,15
0,999	2,3268	2,30
$1-10^{-6}$	3,46	3,45
$1-10^{-10}$	4,59	4,60

On constate que les valeurs arrondies, proportionnelles aux nombres 1, 2, 3, 4, correspondent aux valeurs de Θ(λ) égales à $1-10^{-1}$, $1-10^{-1-2}$, $1-10^{-1-2-3}$, $1-10^{-1-2-3-4}$. On a ainsi un moyen très simple de rétablir l'allure générale de la fonction Θ(λ).

Dans les calculs pratiques, on peut utiliser les logarithmes ordinaires; si l'on pose $e^{-\lambda^2} = 10^{-\mu^2}$, il vient

$$\lambda = \mu\sqrt{\text{Log}\,10},$$

et

$$x = \mu\sqrt{2npq\,\text{Log}\,10}.$$

μ est l'écart relatif décimal, et la courbe $y = \frac{1}{\sqrt{\pi}} e^{-\lambda^2}$ est remplacée par la courbe $y = 10^{-\mu^2}\sqrt{\frac{\text{Log}\,10}{\pi}}$

8. **Loi des grands nombres** — Une conséquence remarquable de la variation de $\Theta(\lambda)$ est que, lorsque n augmente indéfiniment, les parties gagnées et perdues se répartissent suivant les probabilités p et q. C'est cette propriété que l'on appelle « loi des grands nombres »

En effet, u augmente indéfiniment avec n, mais avec l'ordre de croissance de $\sqrt{n}$. Or, si λ est supérieur à 4,6, $\Theta(\lambda)$ est supérieur à $1-10^{-10}$, d'autre part, $\alpha = np + \lambda u$ peut s'écrire

$$\frac{\alpha}{n} = p + \lambda\sqrt{\frac{2pq}{n}}.$$

Donc, lorsque n augmente indéfiniment, la probabilité pour que $\left|\frac{\alpha}{n} - p\right|$ dépasse $5\sqrt{\frac{2pq}{n}}$ est inférieure à 10^{-10}, qui est pratiquement négligeable

D'une manière plus précise, pour que $\left|\frac{\alpha}{n} - p\right|$ ne surpasse pas un nombre donné ε, positif, et aussi petit que l'on veut, il faut et il suffit que l'on ait

$$|\lambda|\sqrt{\frac{2pq}{n}} \leq \varepsilon \quad \text{ou} \quad |\lambda| < \varepsilon\sqrt{\frac{n}{2pq}};$$

la probabilité est donc $\Theta\left(\varepsilon\sqrt{\frac{n}{2pq}}\right)$

On peut toujours donner à n des valeurs suffisamment grandes pour que cette probabilité soit aussi voisine de 1 que l'on veut.

C'est ce que l'on exprime en disant que la probabilité pour que $\left|\frac{\alpha}{n} - p\right|$ soit inférieure à un nombre fixe, positif, ε, aussi petit que l'on veut, tend vers 1 quand le nombre des épreuves augmente indéfiniment.

Ce théorème, appelé encore théorème de Bernoulli, a été généralisé par Tchebychef dans un Mémoire que nous résumons dans la Note n° 1.

II — PROBLÈME DU DEUXIÈME ORDRE

9. Jusqu'ici, le problème que nous avons étudié est un probleme du premier ordre; il était en effet de la nature suivante . étant donnés $n = 2.10^6$ coups de pile ou face, pour lesquels $p = q = \frac{1}{2}$, l'unité d'écart est $u = \sqrt{2npq} = 1000$; le nombre le plus probable des

parties gagnées est 10^6. En posant

$$\alpha = 10^6 + x,$$
$$x = 1000\lambda,$$

α et x ayant la même signification que plus haut, la probabilité pour que $|\lambda|$ soit compris entre λ' et λ'' est $\Theta(\lambda'') - \Theta(\lambda')$.

Mais si nous nous demandons maintenant quelles sont les diverses répartitions possibles des écarts x, pour un très grand nombre N de parties, comportant chacune 2.10^6 coups de pile ou face, nous avons affaire à un problème du deuxième ordre.

On peut se demander tout d'abord quel sera le nombre de parties pour lesquelles l'écart x aura une valeur déterminée ? La probabilité correspondante est égale à $\frac{1}{\sqrt{\pi}} e^{-\lambda^2} d\lambda$, où

$$d\lambda = \frac{dx}{u} = \frac{dx}{1000} = \frac{1}{1000},$$

donc

$$p_x = \frac{1}{1000\sqrt{\pi}} e^{-\frac{x^2}{10^6}}.$$

Par exemple,

$$p_0 = \frac{1}{1000\sqrt{\pi}}; \qquad p_1 = p_{-1} = \frac{1}{1000\sqrt{\pi}} e^{-\frac{1}{10^6}},$$
$$p_2 = p_{-2} = \frac{1}{1000\sqrt{\pi}} e^{-\frac{4}{10^6}}, \quad \ldots$$

Ces probabilités sont représentées par une courbe polygonale qui s'abaisse très lentement jusqu'à $x = \pm 1000$. Mais ceci est la courbe théorique. La courbe pratique présente des écarts par rapport à celle-ci, et ces écarts satisfont eux-mêmes à la loi de Gauss si le nombre N des parties est très grand. Si, par exemple, on fait 5.10^6 parties, et si l'on considère comme événement favorable la nullité de l'écart, dont la possibilité est p_0, l'unité d'écart correspondante est $\sqrt{2Np_0(1-p_0)}$, très sensiblement égale à $\sqrt{2Np_0}$, puisque p_0 est très petit; ce qui fournit, en définitive, une unité d'écart voisine de 100.

Si l'événement favorable est un écart non nul, mais cependant assez faible, l'unité d'écart est encore du même ordre, car les probabilités p_x qui correspondent aux valeurs pas trop grandes de x sont peu différentes.

On sait, par exemple, que l'on a une chance sur deux de dépasser l'écart médian $100 \times 0,4769$, voisin de 50.

En résumé, on peut dire que la courbe pratique comportera des dents de part et d'autre de la courbe théorique, ces dents correspondant à des écarts de l'ordre de 100 pour une ordonnée $Np_0 = \frac{5000}{\sqrt{\pi}}$, sensiblement égale à 3000, l'erreur de la théorie sur les résultats de la pratique apparaît ainsi de l'ordre de $\frac{1}{50}$.

On peut évidemment se proposer le même problème pour une fonction quelconque, c'est-à-dire chercher la répartition des erreurs possibles de la théorie dans le calcul d'une telle fonction, comparé au résultat pratique d'un grand nombre d'expériences.

10. Écart probable. — On appelle écart probable la moyenne des écarts, autrement dit, l'espérance mathématique d'un joueur qui miserait sur l'écart relatif.

Sa valeur est donnée par l'intégrale

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} |\lambda| e^{-\lambda^2} d\lambda = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-\lambda^2} \lambda \, d\lambda = \frac{1}{\sqrt{\pi}} = 0,564.$$

Il ne faut pas confondre l'écart probable avec la valeur la plus probable de l'écart, qui est nulle ; sa valeur est également distincte de celle de l'écart relatif médian. La forme même de la courbe représentative de $\Theta(\lambda)$ rend intuitif le signe de cette différence, puisque l'écart probable est l'abscisse du centre de gravité de l'aire située d'un même côté de l'axe des y, alors que dans le calcul de l'écart relatif médian, toutes les aires interviennent avec le même poids.

11. D'une manière générale, proposons-nous de déterminer la valeur moyenne d'une puissance $p^{\text{ième}}$ de la valeur absolue de l'écart relatif λ (nous dirons aussi, plus brièvement, valeur moyenne d'une puissance $p^{\text{ième}}$ de λ). C'est ce que l'on désigne par « valeur probable » de cette puissance $p^{\text{ième}}$.

Auparavant, il n'est pas inutile de rappeler en quelques lignes les propriétés de l'intégrale eulérienne

$$\Gamma(a) = \int_0^{\infty} e^{-x} x^{a-1} \, dx,$$

qui s'introduit dans le calcul de ces valeurs probables.

Cette intégrale est définie pour a positif, et son intégration par parties conduit immédiatement à la formule de récurrence

$$a\,\Gamma(a) = \Gamma(a+1),$$

qui permet de calculer toutes les valeurs de $\Gamma(a)$, connaissant les valeurs de cette fonction dans l'intervalle $0 < a < 1$.

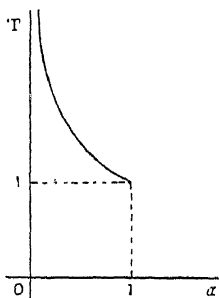



Fig. 8

$\Gamma(0)$ est infini, et, entre 0 et 1, $\Gamma(a)$ décroît constamment de $+\infty$ à 1.

Rappelons encore la relation bien connue [1]

$$\Gamma(a)\,\Gamma(1-a) = \frac{\pi}{\sin a\pi},$$

qui permet de réduire de moitié l'intervalle $(0, 1)$ dans lequel la fonction $\Gamma(a)$ doit être connue. En particulier, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Ce dernier résultat est, d'ailleurs, une conséquence de ce que nous avons déjà vu, car on a

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-x} x^{-\frac{1}{2}}\,dx,$$

ou, en posant $x = y^2$,

$$\Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-y^2}\,dy.$$

En particulier, on peut calculer les valeurs de $\Gamma(a)$ relatives à toutes

---

(1) *Cf.* D'ADHÉMAR, *Leçons sur les principes de l'Analyse*, t. I, Chap. VI.

les valeurs de a multiples de $\frac{1}{2}$. Or, ce sont ces valeurs qui s'introduisent dans le calcul des probabilités; elles peuvent être entières ou fractionnaires.

Par exemple, si a est un nombre entier n,

$$\Gamma(n) = (n-1)!$$

et si a est de la forme $n+\frac{1}{2}$,

$$\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\cdot\left(n-\frac{3}{2}\right)\cdots\frac{3}{2}\,\frac{1}{2}\,\Gamma\left(\frac{1}{2}\right)$$
$$= \frac{(2n-1)\,(2n-3)\ldots 3\,.\,1}{2^n}\sqrt{\pi}$$

Ceci rappelé, déterminons la valeur moyenne d'une puissance $p^{\text{ième}}$ de λ; c'est la somme des produits des valeurs de $|\lambda|^p$ par leurs probabilités, c'est-à-dire

$$\mathfrak{M}(|\lambda|^p) = \frac{2}{\sqrt{\pi}}\int_0^{\infty}\lambda^p e^{-\lambda^2}\,d\lambda$$

Posons $\lambda^2 = z$. Il vient

$$\mathfrak{M}(|\lambda|^p) = \frac{1}{\sqrt{\pi}}\int_0^{\infty} e^{-z} z^{\frac{p-1}{2}}\,dz = \frac{1}{\sqrt{\pi}}\,\Gamma\left(\frac{p+1}{2}\right).$$

Nous désignerons encore cette valeur moyenne par $\overline{|\lambda^p|}$. En particulier,

$$\overline{|\lambda|} = \frac{1}{\sqrt{\pi}},$$

$$\overline{|\lambda^2|} = \frac{1}{2}.$$

La mesure de λ permet ainsi de reconnaître si la suite de parties à laquelle on a affaire est normale ou non.

Si l'on veut les valeurs moyennes des puissances de l'écart absolu x, il suffit de remarquer que dans la formule de définition $x = \lambda u$, u est une constante, égale à $\sqrt{2npq}$, d'où résulte immédiatement

$$\overline{|x^k|} = \frac{2}{\sqrt{\pi}}\int_0^{\infty}|x|^k e^{-\lambda^2}\,d\lambda = u^k\,\overline{|\lambda^k|}.$$

En particulier,

$$\overline{|x|} = \frac{u}{\sqrt{\pi}},$$

$$\overline{|x^2|} = \frac{u^2}{2},$$

si la série de parties est normale, on doit avoir

$$\overline{x^2} = \frac{\pi}{2}\,\overline{|x|}^2,$$

donc la valeur moyenne du carré de l'écart est supérieure au carré de la valeur moyenne de l'écart.

12. *Remarque.* — On sait que la valeur moyenne de x est nulle; celle de $\alpha = np + x$ est, par suite, $\bar{\alpha} = np$.

Si p est petit, $u = \sqrt{2npq} = \sqrt{2q\bar{\alpha}}$ est sensiblement égal à $\sqrt{2\bar{\alpha}}$. Considérons alors une fonction f proportionnelle à α; $f = k\alpha$ ($k > 0$), et supposons que nous déterminions les écarts de f au lieu de ceux de α. Nous avons

$$f = \bar{f} + \xi,$$

en désignant par ξ l'écart de f.

Or

$$\bar{f} = k\bar{\alpha},$$

d'où

$$\xi = kx.$$

Le changement de variable $x = \frac{\xi}{k}$, effectué sur la probabilité

$$\frac{1}{u\sqrt{\pi}}\, e^{-\frac{x^2}{u^2}}\, dx,$$

la met sous la forme

$$\frac{1}{ku\sqrt{\pi}}\, e^{-\frac{\xi^2}{k^2u^2}}\, d\xi$$

La loi de probabilité de f est donc la même que pour α, à condition de prendre la nouvelle unité d'écart $v = ku$.

D'autre part, $v = k\sqrt{2\bar{\alpha}}$ peut s'écrire $v = \sqrt{k}\sqrt{2\bar{f}}$; de sorte que

l'unité d'écart n'est pas $\sqrt{2\bar{j}}$, comme on aurait pu le croire, mais $\sqrt{\bar{\lambda}}$ fois plus grande

Il résulte de cette remarque que la connaissance de $\bar{j}$ ne suffit pas pour en déduire l'unité d'écart correspondante. Il ne suffit pas de savoir que f est proportionnel à σ, il faut encore connaître le facteur de proportionnalité.

13. Nous allons appliquer les résultats précédents à la résolution d'un problème du deuxième ordre.

« Considérons deux phénomènes indépendants, obéissant à la loi normale des écarts. Quelle est la loi de probabilité de la somme de leurs écarts ? »

Soient x et y les écarts de ces deux phénomènes, u et v leurs unités d'écart. Leurs lois respectives de probabilité sont

$$\frac{1}{u\sqrt{\pi}} e^{-\frac{x^2}{u^2}} dx \qquad \text{et} \qquad \frac{1}{v\sqrt{\pi}} e^{-\frac{y^2}{v^2}} dy.$$

Pour que $z = x + y$ soit compris entre deux valeurs infiniment voisines, z et $z + dz$, on peut donner au premier écart une valeur déterminée x; il faut alors que y soit compris entre $z - x$ et $z - x + dz$, et la probabilité pour que ceci ait lieu est

$$\frac{1}{v\sqrt{\pi}} e^{-\frac{(z-x)^2}{v^2}} dz$$

Cette probabilité reste la même tant que x reste compris entre x et $x + dx$, puisque l'infiniment petit dx est arbitraire, et la probabilité correspondante est

$$\frac{1}{u\sqrt{\pi}} e^{-\frac{x^2}{u^2}} dx$$

x pouvant varier de $-\infty$ à $+\infty$, la probabilité cherchée, qui se déduit des deux probabilités précédentes par application du principe de la probabilité composée, a pour valeur

$$\frac{dz}{uv\pi} \int_{-\infty}^{+\infty} e^{-\frac{(z-x)^2}{v^2} - \frac{x^2}{u^2}} dx.$$

L'intégrale est de la forme $\int_{-\infty}^{+\infty} e^{-(ax^2+2bx+c)}dx$, dans laquelle

$$a = \frac{1}{u^2} + \frac{1}{v^2},$$

$$b = -\frac{z}{v^2},$$

$$c = \frac{z^2}{v^2}.$$

En posant $x + \frac{b}{a} = \frac{t}{\sqrt{a}}$, le trinome devient $t^2 + \frac{ac - b^2}{a}$, et l'intégrale s'écrit

$$e^{-\frac{ac-b^2}{a}} \int_{-\infty}^{+\infty} \frac{e^{-t^2}}{\sqrt{a}} dt = \sqrt{\frac{\pi}{a}}\, e^{-\frac{ac-b^2}{a}}.$$

En remplaçant a, b, c par leurs valeurs, nous avons

$$\frac{ac - b^2}{a} = \frac{z^2}{u^2 + v^2},$$

ce qui fournit enfin, pour la probabilité cherchée, l'expression simple

$$\frac{1}{\sqrt{u^2 + v^2}\sqrt{\pi}} e^{-\frac{z^2}{u^2+v^2}} dz,$$

ou encore, en posant $\varpi^2 = u^2 + v^2$,

$$\frac{1}{\varpi\sqrt{\pi}} e^{-\frac{z^2}{\varpi^2}} dz$$

Donc $z = x + y$ suit la loi normale des écarts, avec l'unité d'écart $\varpi = \sqrt{u^2 + v^2}$

Ce résultat remarquable se généralise immédiatement.

Étant données k quantités indépendantes x_1, x_2, ..., x_k, qui suivent la loi normale des écarts, et dont les unités d'écart sont $u_1, u_2, \ldots, u_k$, la quantité $z = a_1x_1 + a_2x_2 + \ldots + a_kx_k$, de coefficients constants, obéit à la même loi, et a pour unité d'écart

$$\varpi = \sqrt{a_1^2u_1^2 + a_2^2u_2^2 + \ldots + a_k^2u_k^2}.$$

Ce résultat n'est plus valable dès que les phénomènes considérés ne sont pas tous indépendants. Par exemple, si x_1 et x_2 sont les

mêmes, $z = (a_1 + a_2)x_1 + a_3x_3 + \ldots + a_kx_k$ obéit bien encore à la loi normale des écarts, mais $w^2 = (a_1 + a_2)^2 u_1^2 + a_3^2 u_3^2 + \ldots + a_k^2 u_k^2$ n'est plus de la forme précédente

14. *Remarque.* — Les opérations autres que l'addition algébrique ne possèdent pas la même propriété.

Considérons, par exemple, deux phénomènes indépendants, obéissant à la loi normale des écarts. Soient x et y leurs écarts, d'unités u et v, et étudions le produit $z = xy$ de ces écarts

x ayant une valeur déterminée, z sera compris entre z et $z + dz$, si y est compris entre $\frac{z}{x}$ et $\frac{z}{x} + \frac{dz}{x}$; la probabilité correspondante est $\frac{1}{v\sqrt{\pi}} e^{-\frac{z^2}{v^2x^2}} \frac{dz}{x}$. Le même raisonnement que pour l'addition conduit alors à la probabilité

$$\frac{dz}{uv\pi}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{u^2}+\frac{z^2}{v^2x^2}}\,dx,$$

qui n'est pas du tout comparable à la loi normale des écarts

Même une fonction aussi simple que $z = x^2$ ne suit pas la loi normale des écarts, en effet, pour que z soit compris entre z et $z + dz$, il faut que x soit compris entre x et $x + \frac{dz}{2x}$, dont la probabilité est

$$\frac{1}{u\sqrt{\pi}} e^{-\frac{z}{u^2}} \frac{dz}{2\sqrt{z}}.$$

15. *Problème* — « On effectue n_1 et n_2 tirages dans deux urnes, contenant chacune des boules blanches et des boules noires. Soient p_1 et p_2 les probabilités respectives de tirer une boule blanche, et, en mettant les écarts en évidence, soient $n_1p_1 + x$ et $n_2p_2 + y$ les nombres de boules blanches tirées.

« En rassemblant ces boules blanches, leur nombre total est

$$n_1p_1 + n_2p_2 + x + y,$$

de valeur moyenne $n_1p_1 + n_2p_2$. Peut-on, avec une seule urne, réaliser le même phénomène? »

Soit p la probabilité de tirer une boule blanche de cette urne, dont on admet l'existence. On doit effectuer $n = n_1 + n_2$ tirages, et la valeur moyenne np doit être égale à

$$np = n_1p_1 + n_2p_2.$$

Il en résulte

$$nq = n_1 q_1 + n_2 q_2.$$

Or, avec les deux urnes, l'écart $z = x + y$ a une unité d'écart w donnée par

$$w^2 = u^2 + v^2 = 2(n_1 p_1 q_1 + n_2 p_2 q_2).$$

Pour l'urne unique, l'unité d'écart W est définie par

$$W^2 = 2npq,$$

ou encore

$$nW^2 = 2np\ nq = 2(n_1 p_1 + n_2 p_2)(n_1 q_1 + n_2 q_2)$$

Formons de même

$$nw^2 = 2(n_1 + n_2)(n_1 p_1 q_1 + n_2 p_2 q_2)$$

La différence de ces deux expressions est alors

$$n(W^2 - w^2) = 2 n_1 n_2 (p_2 - p_1)(q_1 - q_2) = 2 n_1 n_2 (p_2 - p_1)^2 \geqq 0.$$

L'unité d'écart de l'urne unique est donc plus grande que celle fournie par les deux urnes, sauf si $p_2 = p_1$. La réponse est donc négative en général.

Par exemple, si $p_1 = 0$, $q_2 = 0$, u et v, et, par suite, w sont nuls. En outre, si n_1 et n_2 sont proportionnels aux nombres de boules de chaque urne, on réalise la moyenne $n_2 p_2$, relative aux deux urnes, en rassemblant les boules des deux couleurs dans une urne unique, en effet,

$$np = (n_1 + n_2)\frac{n_2}{n_1 + n_2} = n_2 p_2$$

Pourtant, il est clair que, ce faisant, un écart non nul apparaît.

La conclusion est encore négative avec un nombre quelconque d'urnes. Effectuons, d'une manière générale, $n_1, n_2, \ldots, n_k$ tirages dans k urnes, de probabilités $p_1, p_2, \ldots, p_k$.

Si $x_1, x_2, \ldots, x_k$ sont les écarts, d'unités $u_1, u_2, \ldots, u_k$, et si l'on rassemble les boules tirées, l'écart $z = x_1 + x_2 + \ldots + x_k$ a pour unité

$$w = \sqrt{u_1^2 + u_2^2 + \ldots + u_k^2},$$

donc

$$w^2 = 2(n_1 p_1 q_1 + n_2 p_2 q_2 + \ldots + n_k p_k q_k).$$

Réalisons une seule urne, de probabilité p; $n = n_1 + n_2 + \ldots + n_k$

tirages doivent fournir une valeur moyenne

$$np = n_1 p_1 + n_2 p_2 + \dots + n_k p_k,$$

l'unité d'écart est alors $W = \sqrt{2npq}$. Formons encore la différence

$$n(W^2 - w^2) = 2\{(n_1 p_1 + n_2 p_2 + \dots + n_k p_k)(n_1 q_1 + n_2 q_2 + \dots + n_k q_k) - (n_1 + n_2 + \dots + n_k)(n_1 p_1 q_1 + \dots + n_k p_k q_k)\}$$
$$= 2 \sum_{(i,j)} n_i n_j (p_i - p_j)^2 > 0 \qquad (i, j = 1, 2, \dots, k)$$

W et w ne peuvent donc pas être égaux en général.

16. **Théorie de la corrélation.** — Deux phénomènes peuvent présenter entre eux tous les degrés de correspondance entre l'indépendance complete, comme ceux que nous venons d'etudier, et l'identité.

L'étude de la variation, avec ce degre de correspondance, de la valeur moyenne d'une fonction des deux phénomènes, pourra permettre, inversement, de déduire le degre de corrélation des deux phénomenes, de cette valeur moyenne déterminée expérimentalement

Nous allons dire quelques mots de cette theorie de la corrélation, qui a été surtout étudiée et mise en évidence par les météorologistes anglais.

Nous considérerons la valeur moyenne du carré de la somme des écarts $z = x + y$. Si x et y sont indépendants, z suit la loi normale des écarts, avec l'unité d'écart $w = \sqrt{u^2 + v^2}$, nous avons vu que la valeur moyenne de z^2 est alors

$$\overline{z^2} = \frac{w^2}{2} = \overline{x^2} + \overline{y^2}$$

D'autre part

$$z^2 = x^2 + y^2 + 2xy$$

entraîne

$$\overline{z^2} = \overline{x^2} + \overline{y^2} + 2\overline{xy},$$

d'où l'on conclut immédiatement que

$$\overline{xy} = 0.$$

Il fallait s'attendre à ce résultat, car un écart x a autant de proba-

bilité de prendre une valeur positive qu'une valeur négative. On peut considérer $\overline{xy} = 0$ comme un criterium d'indépendance.

Si, au contraire, les deux phénomènes sont identiques,

$$y = x \qquad \text{et} \qquad z = 2x,$$

donc

$$\overline{z^2} = 4\overline{x^2},$$

et

$$2\overline{xy} = \overline{x^2} + \overline{y^2}$$

si

$$y = -x, \qquad 2\overline{xy} = -\overline{x^2} - \overline{y^2}.$$

On est ainsi conduit à considérer le « rapport de corrélation »

$$c = \frac{2\overline{xy}}{\overline{x^2} + \overline{y^2}},$$

et les résultats précédents peuvent être résumés dans le tableau suivant

Si les événements sont concomitants	$c = 1$
Si les événements sont indépendants	$c = 0$
Si les événements sont opposés	$c = -1$

C'est à l'aide de ces formules qu'on cherche à déterminer les corrélations entre les phénomènes divers, par exemple en météorologie, en biologie, etc., à condition que ces phénomènes obéissent à la loi normale des écarts.

17. Voyons comment on peut réaliser différentes corrélations. L'indépendance complète peut être représentée par les tirages de deux personnes différentes dans deux urnes différentes.

D'une manière générale, on peut concevoir trois séries de tirages effectués dans trois urnes distinctes, et d'écarts respectifs t, x_1, y_1. Considérons alors les deux quantités

$$x = t + x_1,$$
$$y = t + y_1.$$

Ces trois tirages étant indépendants, on a

$$\overline{x^2} = \overline{t^2} + \overline{x_1^2},$$
$$\overline{y^2} = \overline{t^2} + \overline{y_1^2}.$$

D'autre part

$$xy = t^2 + tx_1 + ty_1 + x_1y_1,$$

et, par suite,

$$\overline{2xy} = 2\overline{t^2}.$$

Le degré de corrélation de x et y est donc

$$c = \frac{2\overline{t^2}}{2\overline{t^2} + \overline{x_1^2} + \overline{y_1^2}}.$$

Cette représentation ne rend compte que des corrélations positives. On aurait les corrélations négatives en changeant t en $-t$ dans l'expression de y.

Si $t = 0$, on retrouve le cas de deux tirages indépendants, et l'on a bien $c = 0$. Si x_1 et y_1 sont nuls, on retrouve l'identité, et c est égal à 1. Tous les autres cas représentent les corrélations intermédiaires.

Si les trois urnes T, X_1, Y_1 sont identiques, $\overline{t^2}$, $\overline{x_1^2}$, $\overline{y_1^2}$ sont proportionnels aux nombres de tirages effectués dans ces urnes (¹).

Supposons que x corresponde à αN tirages de l'urne T, et $(1-\alpha)$N tirages de l'urne X_1, de même y correspond à αN tirages de T, et $(1-\alpha)$N tirages de Y_1. Alors

$$c = \frac{2\alpha}{2\alpha + 2(1-\alpha)} = \alpha,$$

qui peut prendre toutes les valeurs rationnelles entre 0 et 1, pour un choix convenable des urnes. N'oublions pas que t, x_1, y_1 sont, non les résultats des tirages, mais les écarts de ces résultats par rapport à la moyenne.

18. Corrélation de deux écarts de signes constants. — Lorsque les écarts x et y ont des signes bien déterminés, les corrélations ne sont plus comparables à celles que nous venons de trouver.

On peut évidemment supposer que x et y sont tous deux positifs. Si l'on suppose de plus, pour simplifier l'écriture, que les deux

(¹) Nous verrons que la valeur moyenne du carré de l'écart est proportionnelle au nombre d'épreuves, même pour les lois de probabilité autres que la loi de Gauss. *Cf.* Chap. IV, n° 12.

unités d'écart sont égales à 1, les deux lois de probabilité sont

$$\frac{2}{\sqrt{\pi}} e^{-x^2} dx \qquad \text{et} \qquad \frac{2}{\sqrt{\pi}} e^{-y^2} dy.$$

La corrélation maximum correspond encore à l'identité des deux écarts, et a pour valeur l'unité.

La corrélation moyenne correspond à l'indépendance des deux phénomènes; dans ces conditions,

$$\overline{xy} = \overline{x}\,\overline{y} \qquad \text{et} \qquad c = \frac{2\overline{x}\,\overline{y}}{\overline{x^2}+\overline{y^2}},$$

or

$$\overline{x} = \overline{y} = \frac{1}{\sqrt{\pi}} \qquad \text{et} \qquad \overline{x^2} = \overline{y^2} = \frac{1}{2},$$

ce qui fournit, pour le degré de corrélation, la valeur

$$c = \frac{2}{\pi} = 0,636.$$

Cherchons maintenant le minimum. La question ne pourrait être traitée directement que par la résolution d'un problème difficile du calcul des variations. Nous nous contenterons d'exposer ici un raisonnement élémentaire.

Pour cela nous considérerons la valeur moyenne de xy comme la limite de l'opération élémentaire suivante :

Soient n quantités positives $x_1, x_2, \ldots, x_n$, et n quantités analogues positives $y_1, y_2, \ldots, y_n$, rangées dans l'ordre des valeurs croissantes, l'égalité étant exclue :

$$0 < x_1 < x_2 \ldots x_n,$$
$$0 < y_1 < y_2 \ldots y_n.$$

Effectuons les sommes de la forme

$$S = x_1 y_{i_1} + x_2 y_{i_2} + \ldots + x_n y_{i_n},$$

dans lesquelles les indices $i_1, i_2, \ldots, i_n$ sont les permutations des nombres 1, 2, ..., n, et proposons-nous de déterminer pour quelles permutations de ces indices S est maximum ou minimum.

Pour répondre à cette question, il suffit d'étudier l'effet, sur la valeur de cette somme, de la permutation de deux indices i_1 et i_2.

Soit

$$S' = x_1 y_{i_2} + x_2 y_{i_1} + x_3 y_{i_3} + \ldots + x_n y_{i_n}.$$

La différence de ces deux sommes est

$$S - S' = (x_1 - x_2)(y_{i_1} - y_{i_2})$$

Donc S est supérieur à S' si $y_{i_1} - y_{i_2} < 0$, c'est-à-dire si $i_1 < i_2$, et est, au contraire, inférieur à S' si $i_1 > i_2$. En résumé, on voit que

$$\begin{aligned} S \text{ maximum} &= x_1 y_1 + x_2 y_2 + \ldots + x_n y_n, \\ S \text{ minimum} &= x_1 y_n + x_2 y_{n-1} + \ldots + x_n y_1, \end{aligned}$$

il est remarquable que ces deux sommes extrêmes s'obtiennent d'une façon indépendante de l'ordre de grandeur des y par rapport aux x.

L'application de ce résultat à la détermination de la valeur moyenne minimum de xy nous conduit à multiplier les plus grands écarts x par les plus petits écarts y. On partagera l'intervalle $(0, \infty)$ de x par un certain nombre d'écarts

$$0 < x_1 < x_2 < x_3 < \ldots < x_n,$$

tels que les probabilités, relatives aux intervalles partiels ainsi déterminés, soient égales entre elles, donc égales à $\frac{1}{n+1}$. De même pour y, on considère la suite analogue

$$0 < y_1 < y_2 < y_3 < \ldots < y_n,$$

identique ici à celle des x, et l'on formera

$$x_1 y_n + x_2 y_{n-1} + \ldots + x_n y_1.$$

On voit que ceci revient à multiplier, entre eux, les écarts x et y qui sont liés par la relation

$$\frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx = \frac{2}{\sqrt{\pi}} \int_y^\infty e^{-y^2} dy,$$

ou

$$\Theta(x) = 1 - \Theta(y).$$

Géométriquement, x et y sont tels que les aires hachurées sur la figure représentative de la fonction $\Theta(x)$ soient égales; on conçoit, en effet, d'après la définition même de la probabilité, que des aires égales peuvent être considérées comme contenant le même nombre

d'écarts. On obtient des valeurs approchées satisfaisantes en prenant 10 écarts, très bonnes, avec 100 écarts. Le degré de corrélation minimum est de l'ordre de 0,32, il est donc très appréciable.

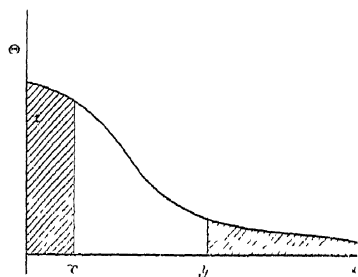



Fig. 9.

Remarquons enfin que la somme maximum

$$S = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

conduit bien à faire le produit de x_i avec l'écart égal y_i, et par suite à la limite, à prendre $y = x$.

CHAPITRE IV.

PROBABILITÉS CONTINUES PROBLÈMES DU DEUXIÈME ORDRE.

I — PROBLÈMES SUR LA DROITE

1. Dans ce Chapitre, nous nous proposons d'étudier un problème analogue à celui que nous venons de résoudre sur les probabilités discontinues, et qui va nous conduire à une loi différente, mais également très générale et très importante.

Ce problème est relatif à la distribution de points sur une droite. Nous commencerons par examiner le cas d'un nombre fini de points sur un segment de droite, et nous passerons ensuite au cas des infiniment grands

Problème. — Soit un segment de droite OA de longueur l On suppose qu'un point M, placé sur ce segment, a une probabilité égale de se trouver en un point quelconque du segment, autrement dit, sa probabilité de tomber sur un élément dx de OA est $\frac{dx}{l}$, quelle que soit la situation de dx sur le segment OA En particulier, la probabilité pour que M soit sur OA est une certitude.

« Dans ces conditions, plaçons n points au hasard. L'un d'entre eux, M', est plus près de O que tous les autres. Quelle est sa distance probable à O ? »

Pour résoudre cette question, il faut tout d'abord établir des lois de probabilité. Déterminons la probabilité pour qu'il n'y ait aucun des n points sur un segment OB de longueur $y < l$.

Il faut d'abord que le premier point M_1 n'y soit pas, événement dont la probabilité est $\left(1 - \frac{y}{l}\right)$; et de même pour chacun des $(n-1)$ points suivants. Par suite de l'indépendance de ces événements, la probabilité cherchée est $\left(1 - \frac{y}{l}\right)^n$

Plus loin, nous allons faire croître l et n indéfiniment. Le résultat précédent dépendra alors de l'ordre de grandeur de ces deux quantités. Il est naturel de supposer que la « densité linéaire » $\delta = \frac{n}{l}$ reste la même (ou, ce qui revient au même, a une limite). C'est ce que nous ferons.

Dans ces conditions la probabilité que nous venons d'établir s'écrit $\left(1 - \frac{x}{l}\right)^{l\delta}$. Si maintenant l augmente indéfiniment, δ restant fixe, cette expression a une limite bien déterminée $e^{-\delta x}$.

En résumé, la probabilité pour que le point le plus rapproché de O, parmi les points placés sur une demi-droite OX avec une densité linéaire donnée δ, soit à une distance de O supérieure à x, est $e^{-\delta x}$.

Il en résulte immédiatement que la probabilité pour que le point le plus rapproché soit compris entre x et $x + dx$ est

$$\varphi(x)\,dx = -d(e^{-\delta x}) = \delta\, e^{-\delta x}\,dx,$$

d'où

$$\varphi(x) = \delta\, e^{-\delta x}.$$

Enfin, il est bien clair que la probabilité pour qu'il n'y ait aucun point sur un segment quelconque x de OX est également $e^{-\delta x}$.

2. La même question peut être reprise à un point de vue un peu différent, qui nous permettra de l'étudier d'une manière plus approfondie. Considérons le même segment OA de longueur l, sur lequel nous plaçons N points ; leur densité linéaire est $\delta = \frac{N}{l}$.

« Étant donné un segment x de OA, quelle est la probabilité pour que n des N points soient sur x ? »

Cette probabilité p_n est liée à la probabilité plus particulière ϖ_n pour que se trouvent, sur x, n points donnés d'avance (par un numéro dont on affecte chaque point).

On obtient évidemment p_n en multipliant ϖ_n par le nombre des combinaisons n à n des N points numérotés :

$$p_n = \varpi_n \frac{N!}{n!\,(N-n)!}.$$

Pour calculer ϖ_n, formons le rapport $\frac{\varpi_{n+1}}{\varpi_n}$. Soient $A_1, A_2, \ldots, A_n$ les n points situés sur x, dans ϖ_n, et $A_{n+1}, \ldots, A_N$ ceux qui sont répartis sur la longueur restante $l-x$. A chaque probabilité ϖ_n, correspond une probabilité ϖ_{n+1}, dans laquelle ce sont les points $A_1, A_2, \ldots, A_n, A_{n+1}$ qui sont situés sur x.

Le rapport $\frac{\varpi_{n+1}}{\varpi_n}$ de ces deux probabilités est donc égal au rapport des probabilités pour le point A_{n+1}, soit sur x et sur $l-x$, c'est-à-dire

$$\frac{\varpi_{n+1}}{\varpi_n} = \frac{x}{l-x}.$$

On a donc

$$\frac{p_{n+1}}{p_n} = \frac{\varpi_{n+1}}{\varpi_n}\,\frac{N-n}{n+1} = \frac{x}{l-x}\,\frac{N-n}{n+1}.$$

Supposons maintenant que N et l augmentent indéfiniment, la densité linéaire $\delta = \frac{N}{l}$ restant constante. A la limite, il vient

$$\frac{p_{n+1}}{p_n} = \frac{\delta x}{n+1}.$$

$\nu = \delta x$ représente le nombre probable de points sur le segment x, et il est clair que ce n'est pas un nombre entier en général

Cette formule de récurrence permet de former le tableau

$$p_1 = \frac{\nu}{1}p_0,$$

$$p_2 = \frac{\nu^2}{1.2}p_0,$$

$$\ldots\ldots,$$

$$p_n = \frac{\nu^n}{n!}p_0,$$

$$\ldots\ldots$$

Nous avons établi d'autre part que $p_0 = e^{-\nu}$, donc

$$p_0+p_1+\ldots+p_n+\ldots = e^{-\nu}\left(1+\frac{\nu}{1}+\frac{\nu^2}{1.2}+\ldots+\frac{\nu^n}{n!}+\ldots\right) = 1.$$

On peut dire que la probabilité p_∞ pour qu'il y ait une infinité de points sur x est nulle [1].

---

[1] La signification précise de cet énoncé a été établie précédemment au sujet des probabilités dénombrables. *Cf.* Chap. II, § 11

Ce résultat paraît d'ailleurs évident, puisque la densité linéaire δ est finie.

Si x est très petit, et si nous le désignons par dx, nous avons $p_0 = 1 - \delta\, dx$, $p_1 = \delta\, dx$, et les probabilités suivantes sont des infiniment petits d'ordres supérieurs au premier.

Ces résultats permettent de répondre immédiatement à la question que nous nous sommes posée initialement pour une demi-droite OX. On peut, en effet, considérer la probabilité $\varphi(x)\, dx$ pour que le point le plus à gauche soit à une distance de O comprise entre x et

$x + dx$, comme composée de la probabilité pour qu'il n'y ait aucun point sur le segment x, et de la probabilité pour qu'il y ait un point dans dx. Ces deux probabilités étant indépendantes, la probabilité cherchée est leur produit

$$\varphi(x)\, dx = e^{-\delta x} \delta\, dx.$$

3. Longueur moyenne d'un segment. — Supposons que les points placés sur la demi-droite OX, avec la densité linéaire δ, soient numérotés, et soient A_1, A_p, A_q, A_r, ... ces points, répartis de gauche à droite, à partir de l'extrémité O. Posons

$$x_1 = A_1 A_p, \quad x_2 = A_p A_q, \quad x_3 = A_q A_r, \quad \ldots, \quad x_i = A_s A_t, \quad \ldots$$

« Quelle est la probabilité pour que le $i^{\text{ème}}$ segment ait une longueur x_i comprise entre x et $x + dx$? »

Il est clair qu'on ne change pas la densité linéaire en négligeant les points situés à gauche de x_i, et en faisant jouer à A_s le rôle que jouait O dans le problème initial. Dans ces conditions, il faut que le point A_t, le plus rapproché de A_s, soit à une distance comprise entre x et $x + dx$, et nous savons que la probabilité de cet événement est $e^{-\delta x} \delta\, dx$.

Il résulte de là que, sur m segments considérés, le nombre probable de ceux dont la longueur est comprise entre x et $x + dx$ est $m\, e^{-\delta x} \delta\, dx$. La longueur moyenne de ces m segments est donc

$$\frac{1}{m} \int_0^\infty x\, m\, e^{-\delta x} \delta\, dx = \int_0^\infty \delta x\, e^{-\delta x}\, dx,$$

ou encore, en posant $\delta x = y$,

$$\frac{1}{\delta}\int_0^\infty e^{-y} y \, dy = \frac{1}{\delta}\Gamma(2) = \frac{1}{\delta}.$$

La longueur moyenne des segments constitués par deux points consécutifs est donc l'inverse de la densité linéaire.

Remarque. — On peut avoir une conception différente de la longueur moyenne. Imaginons, en effet, un observateur qui se place au hasard sur OX. Il tombe à l'intérieur d'un segment x_i, dont il prend la mesure. S'il recommence plusieurs fois cette expérience, il appellera longueur moyenne la moyenne des longueurs qu'il aura mesurées. Il est évident que cette définition ne concorde pas avec la précédente, car, ici, un segment sera mesuré d'autant plus souvent qu'il sera plus grand.

Sur m segments, la longueur totale de ceux dont la longueur est comprise entre x et $x + dx$ est

$$m\,\delta x\, e^{-\delta x}\, dx,$$

et le rapport de cette longueur à la longueur totale $\frac{m}{\delta}$ des m segments est

$$\delta^2 x\, e^{-\delta x}\, dx.$$

C'est ce rapport qui mesure la probabilité pour que l'observateur tombe dans un segment de longueur comprise entre x et $x + dx$. S'il fait A expériences en tout, la longueur x sera mesurée $A\,\delta^2 x\, e^{-\delta x}\, dx$ fois, et la longueur moyenne cherchée est

$$\frac{1}{A}\int_0^\infty A\,\delta^2 x^2 e^{-\delta x}\, dx = \int_0^\infty \delta^2 x^2 e^{-\delta x}\, dx,$$

ou, en posant encore $y = \delta x$,

$$\frac{1}{\delta}\int_0^\infty e^{-y} y^2\, dy = \frac{1}{\delta}\Gamma(3) = \frac{2}{\delta}.$$

C'est exactement le double de la longueur moyenne à laquelle nous avait conduits la première définition.

4. Avant de passer à l'étude de la probabilité $p_n = e^{-\nu}\frac{\nu^n}{n!}$, remar-

quons qu'elle se rattache étroitement au problème des probabilités discontinues étudié dans le Chapitre précédent.

Pour raisonner à ce point de vue, nous partagerons le segment x en m parties égales, de longueur $\frac{x}{m}$. On peut évidemment supposer m assez grand pour qu'il n'y ait qu'un des n points, au maximum, sur chacun de ces segments partiels.

La densité linéaire étant toujours désignée par δ, l'espérance mathématique relative à un segment $\frac{x}{m}$ est $p = \frac{\delta x}{m} = \frac{\nu}{m}$, de sorte que $q = 1 - \frac{\nu}{m}$.

La probabilité d'avoir n points sur x est alors celle de réaliser n cas favorables de probabilité p, sur les m expériences que représentent ces m segments $\frac{x}{m}$; son expression, bien connue, est

$$p^n q^{m-n} \frac{m!}{n!(m-n)!} = \frac{\nu^n}{n!}\left(1 - \frac{\nu}{m}\right)^{m-n} \frac{m!}{m^n(m-n)!}.$$

Pour revenir à la continuité, considérée comme limite du discontinu, il suffit de faire croître m indéfiniment; la fraction $\frac{m!}{m^n(m-n)!}$ tend vers 1, et l'on retrouve bien p_n.

5. Étude de $p_n = e^{-\nu}\frac{\nu^n}{n!}$. — L'expression de p_n est maximum lorsque $\frac{p_n}{p_{n-1}} = \frac{\nu}{n}$ passe par la valeur 1, c'est-à-dire pour $n = \nu$ et $n = \nu - 1$, en supposant que ν soit un nombre entier; ce maximum est alors

$$p_{\nu-1} = p_\nu = \frac{e^{-\nu}\nu^\nu}{\nu!}.$$

Si ν est très grand, p_ν est de la forme

$$p_\nu = \frac{1 - \varepsilon_\nu}{\sqrt{2\pi\nu}},$$

ε_ν étant équivalent à $\frac{1}{12\nu}$.

Pour les autres valeurs de n, il est naturel de poser $n = \nu + h$; on a alors

$$p_n = \frac{e^{-\nu}\nu^{\nu+h}}{(\nu+h)!}.$$

Si h est infiniment petit par rapport à l'infiniment grand ν, $(\nu+h)!$ est équivalent à

$$e^{-\nu-h}(\nu+h)^{\nu+h+\frac{1}{2}}\sqrt{2\pi},$$

d'où

$$p_n = \frac{e^h}{\left(1+\frac{h}{\nu}\right)^{\nu+h+\frac{1}{2}}} \frac{1}{\sqrt{2\pi\nu}} = \frac{e^{-\Lambda}}{\sqrt{2\pi\nu}},$$

en posant

$$\Lambda = \left(\nu+h+\frac{1}{2}\right)\log\left(1+\frac{h}{\nu}\right) - h = \frac{1}{2}\frac{h^2}{\nu} + \frac{h}{2\nu} = \frac{\left(h+\frac{1}{2}\right)^2}{2\nu} - \frac{1}{8\nu}.$$

Si nous posons enfin $\dfrac{h+\frac{1}{2}}{\sqrt{2\nu}} = \lambda$, il vient

$$p_n = e^{\frac{1}{8\nu}} \frac{e^{-\lambda^2}}{\sqrt{2\pi\nu}}.$$

On retrouve ainsi la loi normale des écarts, dans laquelle l'écart est $h+\frac{1}{2} = n - \left(\nu - \frac{1}{2}\right)$, et l'unité d'écart $\sqrt{2\nu}$, λ est alors l'écart relatif. D'ailleurs, si l'on revient aux probabilités discontinues en partageant le segment x en un grand nombre m de parties égales, on vérifie bien que $\sqrt{2mpq}$ est équivalent à $\sqrt{2\nu}$, car $p = \frac{\nu}{m}$ est infiniment petit, et, par suite, q très voisin de 1

6. **Valeurs moyennes** — Une succession d'expériences de repartition de points sur la demi-droite OX fournit, pour une fonction de ces points, une succession de valeurs, dont on est conduit, comme dans les problèmes de probabilités discontinues, à rechercher la moyenne; cette valeur moyenne caractérisera une suite normale d'expériences.

Désignons par $\overline{n}$, $\overline{n^2}$, $\overline{n^3}$, ... les valeurs moyennes des puissances successives de n, n étant le nombre de points placés sur un segment x de la demi-droite. Par définition, nous avons

$$\overline{n} = \sum_{n=1}^{\infty} np_n = e^{-\nu}\sum_{n=1}^{\infty}\frac{\nu^n}{(n-1)!} = \nu$$

Ainsi, ce que nous avons désigné par h est l'écart entre n et sa

valeur moyenne. Il y a donc un certain intervalle entre cette valeur moyenne et l'axe de symétrie $n = \nu - \frac{1}{2}$ de la courbe des probabilités. Nous retrouvons ici une remarque déjà faite à propos des probabilités discontinues.

Au lieu de calculer directement $\overline{n^2}$, $\overline{n^3}$, ..., il est plus simple, et l'on comprendra tout de suite pourquoi, de passer par l'intermédiaire des valeurs moyennes de $n(n-1)$, $n(n-1)(n-2)$, ...

On a

$$\overline{n(n-1)} = \sum_{n=2}^{\infty} n(n-1)p_n = e^{-\nu}\sum_{n=2}^{\infty}\frac{\nu^n}{(n-2)!} = \nu^2,$$

donc

$$\overline{n^2} = \overline{n(n-1)} + \overline{n} = \nu^2 + \nu,$$

de même,

$$\overline{n(n-1)(n-2)} = \nu^3,$$

et, d'autre part,

$$\overline{n(n-1)(n-2)} = \overline{n^3} - 3\overline{n^2} + 2\overline{n},$$

d'où l'on déduit

$$\overline{n^3} = \nu^3 + 3\nu^2 + \nu,$$

et ainsi de suite.

Applications. — Une application immédiate de ces résultats est la détermination du nombre moyen des segments formés avec les points situés sur le segment x considéré.

Les n points répartis sur ce segment définissent $\frac{n(n-1)}{2}$ segments, dont le nombre moyen est

$$\frac{\overline{n(n-1)}}{2} = \frac{\nu^2}{2} = \frac{\delta^2 x^2}{2}.$$

Étant donnés deux segments x, x', sans partie commune, le nombre moyen des segments qui sont à cheval sur x et x' est donc

$$\overline{\left[\frac{(n+n')(n+n'-1)}{2} - \frac{n(n-1)}{2} - \frac{n'(n'-1)}{2}\right]} = \overline{nn'}$$

et

$$\overline{nn'} = \nu\nu',$$

puisque les deux segments sont distincts.

On peut encore obtenir ce résultat en raisonnant directement sur

les valeurs moyennes. Le nombre moyen des segments dont les extrémités sont sur l'ensemble $x + x'$ est $\frac{\delta^2 (x + x')^2}{2}$, donc le nombre moyen cherché est

$$\frac{\delta^2(x + x')^2}{2} - \frac{\delta^2 x^2}{2} - \frac{\delta^2 x'^2}{2} = \delta^2 x x' = \nu\nu'$$

« En résumé, le nombre moyen des segments qui chevauchent sur deux segments distincts donnés est égal au produit des nombres moyens de points sur chacun de ces deux segments. »

Remarquons enfin que ces résultats ne supposent aucune hypothèse sur l'ordre de la grandeur de ν, car nous avons pris p_n sous sa forme générale. Il en sera de même dans le paragraphe suivant.

7. Proposons-nous maintenant de déterminer les valeurs probables de h et de ses puissances. En posant toujours $n = \nu + h$, on a

$$\overline{n} = \nu + \overline{h},$$

donc $|\overline{h}| = o$.

Remarquons, à ce sujet, que la loi normale

$$p_n = e^{\frac{1}{8\nu}} \frac{e^{-\lambda^2}}{\sqrt{2\pi\nu}},$$

qui n'est valable que pour les valeurs très grandes de ν, conduirait à la valeur moyenne $-\frac{1}{2}$.

De même,

$$\overline{n^2} = \overline{(\nu + h)^2} = \nu^2 + 2\nu\overline{h} + \overline{h^2} = \nu^2 + \nu,$$

d'où

$$\overline{h^2} = \nu.$$

On en déduit

$$\overline{\lambda^2} = \frac{1}{2\nu}\left(\overline{h^2} + \overline{h} + \frac{1}{4}\right) = \frac{1}{2} + \frac{1}{8\nu},$$

et, lorsque ν augmente indéfiniment, cette valeur moyenne tend bien vers la valeur $\frac{1}{2}$, que nous avions trouvée pour les probabilités discontinues.

La valeur moyenne de $|h|$ s'obtient par un calcul direct:

$$\overline{|h|} = \sum_{n=0}^{\nu} (\nu - n) p_n + \sum_{n=\nu}^{\infty} (n - \nu) p_n.$$

Or

$$np_n = \nu p_{n-1},$$

d'où résulte

$$\sum_{n=0}^{\nu}(\nu - n)p_n = \nu\left\{\sum_{n=0}^{\nu}p_n - \sum_{n=0}^{\nu-1}p_n\right\} = \nu p_\nu$$

et

$$\sum_{n=\nu}^{\infty}(n-\nu)p_n = \nu\left\{\sum_{n=\nu-1}^{\infty}p_n - \sum_{n=\nu}^{\infty}p_n\right\} = \nu p_{\nu-1} - \nu p_\nu,$$

et l'on a enfin

$$\overline{|h|} = 2\nu p_\nu.$$

Cette valeur dépend donc de ν. Si ν est très grand, elle est équivalente à

$$\frac{2\nu}{\sqrt{2\pi\nu}} = \sqrt{\frac{2\nu}{\pi}}.$$

On peut alors écrire

$$\frac{\overline{h^2}}{\overline{|h|}^2} = \frac{\pi}{2}.$$

Donc, si ν est très grand, on retrouve bien le même rapport que pour la loi normale des écarts; au contraire, si ν est fini, ce rapport $\frac{\overline{h^2}}{\overline{|h|}^2}$ peut être complètement différent de $\frac{\pi}{2}$. Par exemple, pour $\nu = 1$, on a encore $\overline{h^2} = \nu = 1$, mais, d'autre part, $\overline{|h|} = \frac{2}{e}$, ce qui donne pour ce rapport la valeur $\frac{e^2}{4}$.

8. Étude des couples de points. — Les problèmes que nous venons de traiter sont susceptibles d'une double généralisation. On peut raisonner sur des groupes de points répartis sur une droite, au lieu de raisonner sur ces points pris isolément, ou encore transporter ces problèmes dans un espace à plusieurs dimensions.

Nous ne nous occuperons, pour l'instant, que de la première généralisation. Nous supposons toujours que des points sont répartis sur une droite avec une densité linéaire δ, et nous allons étudier les couples de points situés sur un segment x.

« Soit A, B un tel couple. La distance AB est inférieure à x; quelle est sa probabilité pour qu'elle soit inférieure à une longueur donnée ε, elle-même inférieure ou égale à x ? »

Ce problème n'est pas nouveau, et nous l'avons déjà résolu comme premier exemple de probabilité continue, et nous avions trouvé, pour que AB soit compris entre γ et $\gamma + d\gamma$, la probabilité

$$\theta(\gamma)\, d\gamma = \frac{2}{x}\left(1 - \frac{\gamma}{x}\right) d\gamma.$$

Donc la probabilité pour que AB soit inférieur à ε est

$$P(\varepsilon) = \int_0^{\varepsilon} \theta(\gamma)\, d\gamma = \frac{2\varepsilon}{x} - \frac{\varepsilon^2}{x^2}.$$

On peut également arriver à ce résultat en introduisant une deuxième dimension. Pour cela, menons par une extrémité O du segment x, un segment perpendiculaire et de même longueur. On choisira A sur le premier segment, et l'on reportera B sur le deuxième segment, en B', à la même distance de O que B. Le couple AB peut être ainsi représenté par le sommet M du rectangle construit sur ces deux segments OA et OB'. La distance des deux points A, B est mesurée par la longueur MN.

Donc pour que AB soit inférieur à ε, il faut et il suffit que M soit dans la bande définie par $|\xi - \eta| \leq \varepsilon$, et la probabilité cherchée $P(\varepsilon)$

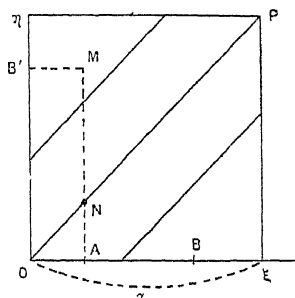



Fig. 10.

est le rapport de l'aire de cette bande à l'aire du carré, rapport dont le calcul est immédiat.

« Supposons maintenant que, dans notre répartition de points, n de ceux-ci se trouvent sur le segment x. Quel est le nombre probable de distances, telles que AB, qui ne sont pas supérieures à ε ? »

Ce nombre représente l'espérance mathématique d'un joueur qui gagnerait 1 pour chaque couple dont la distance est inférieure ou égale à ε. Ajouter les espérances relatives à chaque couple, pris isolément, revient à considérer ces couples comme indépendants, or, ils ne le sont pas, car, par exemple, si A_1A_2 est inférieur à ε, ainsi que A_2A_3, A_1A_3 a, en général, une probabilité plus grande d'être inférieur à ε que si l'on prend A_1 et A_3 arbitrairement. Mais il faut remarquer que le joueur mise d'avance sur tous les couples; il peut même vendre séparément les espérances mathématiques relatives à ces différents couples [1]. A ce point de vue, les espérances s'additionnent, et l'espérance totale est égale au produit de $P(\varepsilon)$ par le nombre des segments; c'est donc

$$\frac{n(n-1)}{2}\left(\frac{2\varepsilon}{x}-\frac{\varepsilon^2}{x^2}\right).$$

Dans une succession d'expériences, le nombre moyen sera $\overline{\frac{n(n-1)}{2}}\left(\frac{2\varepsilon}{x}-\frac{\varepsilon^2}{x^2}\right)$, c'est-à-dire

$$\delta^2\varepsilon x-\frac{\delta^2\varepsilon^2}{2}.$$

Si ε est petit par rapport à x, $\frac{\delta^2\varepsilon^2}{2}$ est négligeable devant $\delta^2\varepsilon x$. Nous verrons d'ailleurs que ce terme $\frac{\delta^2\varepsilon^2}{2}$ est dû à l'influence des extrémités du segment x, influence d'autant plus négligeable que x est plus étendu.

9. La densité des segments dont la longueur est inférieure à ε est

$$\frac{1}{x}\left(\delta^2\varepsilon x-\frac{\delta^2\varepsilon^2}{2}\right)=\delta^2\varepsilon-\frac{\delta^2\varepsilon^2}{2x}.$$

Cette densité est équivalente à $\delta^2\varepsilon$ lorsque ε est infiniment petit par rapport à x. Ce dernier résultat peut s'obtenir facilement en faisant appel à la notion de « densité des extrémités » de ces segments.

Par hypothèse, on néglige l'influence des extrémités de x. Étant alors donné un point A_1, portons la longueur ε de part et d'autre de

[1] Ce genre de raisonnement n'est pas nouveau. Nous l'avons déjà employé dans le problème des rencontres et dans le problème de l'aiguille.

ce point Il faut que le deuxième point A_2 soit sur le segment 2ε ainsi obtenu, si l'on veut que A_1A_2 soit inférieur ou égal à ε. Le nombre moyen des points situés sur ce segment étant $\delta . 2\varepsilon$, la densité, pour le segment total x, des extrémités des segments inférieurs à ε est $\delta . \delta . 2\varepsilon = 2\varepsilon\delta^2$, qui est bien le double de la densité de ces segments

Remarquons que, dans ce raisonnement, on néglige l'existence *a priori* de A_1, dans le segment 2ε qui admet ce point pour milieu Mais il est facile de voir que, malgré cette restriction, notre raisonnement est rigoureux, car il conduit à prendre pour densité des points A_2 du segment 2ε, $\delta = \frac{N}{l}$, alors qu'on devrait prendre $\frac{N-1}{l}$ (N et l ayant la même signification qu'au début de ce Chapitre); or ces deux quantités sont équivalentes lorsque N et l sont infiniment grands

Le résultat que nous venons d'obtenir conduit immédiatement à une conséquence intéressante Soit, en effet, une longueur ε du segment x. Sur cette longueur, le nombre moyen des points qui sont l'extrémité d'un segment inférieur à ε est $2\varepsilon^2\delta^2$.

Parmi ces segments, certains ont leurs deux extrémités sur ε, les autres n'en ont qu'une Or on sait que le nombre moyen des extrémités des premiers est $\overline{n(n-1)} = \delta^2\varepsilon^2$, donc les seconds sont également en nombre $\delta^2\varepsilon^2$

« En résumé, sur un segment de longueur ε, il y a, en moyenne, autant de segments intérieurs que de segments inférieurs à ε qui empiètent. »

10. Portion utile moyenne. — Lorsque l'on considère un point arbitraire M du segment x, ce point peut être l'extrémité d'un segment MN de longueur inférieure à ε, si N se trouve dans la partie commune à x et au segment 2ε de milieu M. C'est cette partie commune que l'on désigne sous le nom de « portion utile ». La moyenne de cette portion utile peut se calculer par la méthode qui nous a déjà fourni la probabilité pour que MN soit inférieur à ε. Mais on peut également la déduire de $P(\varepsilon)$

Si ε' est cette valeur moyenne, la probabilité en question est $\frac{\varepsilon'}{x}$, ce qui donne

$$\frac{\varepsilon'}{x} = P(\varepsilon),$$

d'où

$$\varepsilon' = 2\varepsilon - \frac{\varepsilon^2}{x}.$$

Considérons alors les n points $M_1, M_2, \ldots, M_n$ répartis sur x, et supposons que ces points se placent l'un après l'autre. Quand on place M_1, on fait apparaître une portion utile ε', utilisable par M_2; quand on place ensuite M_2, on fait apparaître une nouvelle portion utile ε', et M_3 donnera un segment inférieur à ε s'il se trouve sur l'une ou l'autre de ces deux portions utiles. Si elles ont une partie commune, un point de cette partie donnera deux segments, l'un avec M_1, l'autre avec M_2, de sorte que, dans tous les cas, la portion utilisable par M_3 est $2\varepsilon'$. Le même raisonnement montrerait que la partie utilisable par M_4 est $3\varepsilon'$, et ainsi de suite, jusqu'au point M_n, qui peut utiliser une longueur $(n-1)\varepsilon'$.

Remarquons que, dans ce raisonnement, chaque segment n'est compté qu'une fois, avec son extrémité d'indice le plus élevé; en définitive, l'espérance totale d'un joueur qui gagnerait 1 franc pour chacun des segments de longueur inférieure à ε est

$$\frac{\varepsilon'}{x}[1+2+3+\ldots+(n-1)] = \frac{n(n-1)}{2}\frac{\varepsilon'}{x} = \frac{n(n-1)}{2} p_\varepsilon.$$

11. Valeurs moyennes relatives aux couples de points. — Nous savons que le nombre moyen des segments de longueur inférieure à ε est $\frac{\nu^2}{2}\left(\frac{2\varepsilon}{x} - \frac{\varepsilon^2}{x^2}\right)$, où ν a sa signification habituelle. Si ε est égal à x, on retrouve bien $\frac{\nu^2}{2}$, et l'écart est la différence $\frac{n(n-1)}{2} - \frac{\nu^2}{2}$, dont la valeur moyenne est zéro.

Calculons les valeurs moyennes de la valeur absolue et du carré de cet écart. La probabilité d'avoir n points sur x est $p_n = e^{-\nu}\frac{\nu^n}{n!}$, ce qui donne

$$\overline{\left|\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right|} = \sum_{n=0}^{\nu'}\left[\frac{\nu^2}{2} - \frac{n(n-1)}{2}\right]p_n + \sum_{n=\nu'+1}^{\infty}\left[\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right]p_n,$$

où ν' est l'entier qui vérifie la double inégalité

$$\nu'(\nu'-1) < \nu^2 \leq \nu'(\nu'+1).$$

En particulier, $\nu' = \nu$ si ν est un nombre entier.

Or $n(n-1)p_n = \nu^2 p_{n-2}$, de sorte que

$$\sum_{n=0}^{\nu'}\left[\frac{\nu^2}{2} - \frac{n(n-1)}{2}\right]p_n = \frac{\nu^2}{2}(p_{\nu'} + p_{\nu'-1})$$

et

$$\sum_{n=\nu'+1}^{\infty}\left[\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right]p_n = \frac{\nu^2}{2}(p_{\nu'-1} + p_{\nu'});$$

il vient enfin

$$\overline{\left|\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right|} = \nu^2(p_{\nu} + p_{\nu'-1})$$

Si ν est entier, et infiniment grand, $p_\nu = p_{\nu-1} = \frac{1}{\sqrt{2\pi\nu}}$, et la valeur moyenne de la valeur absolue de l'écart est

$$\frac{2\nu^2}{\sqrt{2\pi\nu}} = \nu\sqrt{\frac{2\nu}{\pi}}.$$

Pour la valeur moyenne du carré, on a

$$\overline{\left[\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right]^2} = \frac{1}{4}\left[\overline{n^2(n-1)^2} - 2\nu^2\,\overline{n(n-1)} + \nu^4\right].$$

Or

$$n^2(n-1)^2 = n(n-1)(n-2)(n-3) + n(n-1)(4n-6)$$

et

$$n(n-1)(4n-6) = 4n(n-1)(n-2) + 2n(n-1),$$

donc

$$\begin{aligned}\overline{n^2(n-1)^2} &= \overline{n(n-1)(n-2)(n-3)} + 4\,\overline{n(n-1)(n-2)} + 2\,\overline{n(n-1)}\\ &= \nu^4 + 4\nu^3 + 2\nu^2\end{aligned}$$

D'où résulte enfin la valeur moyenne

$$\overline{\left[\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right]^2} = \nu^2\left(\nu + \frac{1}{2}\right).$$

Remarquons que, lorsque ν augmente indéfiniment, le carré moyen est de l'ordre de ν^3, c'est-à-dire de l'ordre du carré de la fluctuation moyenne, et le rapport $\dfrac{\overline{\left[\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right]^2}}{\overline{\left|\frac{n(n-1)}{2} - \frac{\nu^2}{2}\right|}^2}$ tend vers $\frac{\pi}{2}$. Ce résultat

vérifie l'analogie, sur laquelle nous avons déjà insisté, avec la loi normale des écarts, lorsque ν est infiniment grand.

Remarque. — Si x est égal à $h\varepsilon$, h étant un nombre entier, le nombre moyen des segments inférieurs à ε est $\frac{2h-1}{h^2}\frac{\nu^2}{2}$; or

$$\nu = \delta x = h\,\delta\varepsilon,$$

donc ce nombre moyen s'écrit encore $(2h-1)\frac{\delta^2\varepsilon^2}{2}$; il est donc $2h-1$ fois plus grand que pour le segment ε. On comprend pourquoi ces segments se rangent en deux catégories ; il y en a $h\frac{\delta^2\varepsilon^2}{2}$ qui sont intérieurs à l'un des h segments ε en lesquels on peut diviser x, et $(h-1)\frac{\delta^2\varepsilon^2}{2}$ qui chevauchent sur l'un quelconque des $(h-1)$ couples de segments ε consécutifs.

12. La fluctuation du carré de l'écart est d'emploi plus pratique que la fluctuation de cet écart lui-même. Nous allons voir également que son emploi a un caractère plus général.

Dans ce qui précède, il s'agissait d'une densité linéaire δ qui était le résultat d'une infinité d'expériences, de sorte que la probabilité d'amener un nombre n de points sur un segment donné x n'était jamais nulle, quelque grand que fût n. Plaçons-nous maintenant dans le cas d'un nombre limité d'expériences, le jeu de dé par exemple.

Dans une partie de 6 coups, le n° 6 sera amené 1 fois, en moyenne ; les répartitions possibles sont 0, 1, 2, ... ou 6 fois, qui ne sont pas symétriques par rapport au nombre moyen 1.

Plus généralement, considérons N expériences, dans lesquelles la probabilité du cas favorable est p, et celle du cas défavorable $q = 1 - p$. On sait que la probabilité d'amener n fois l'événement favorable, dans ces N expériences, est

$$P_n = \frac{N(N-1)\dots(N-n+1)}{n!} p^n q^{N-n}.$$

Le nombre moyen est

$$\bar{n} = \sum_{n=1}^{N} n P_n = Np,$$

ce que nous savions déjà. De même, la valeur moyenne de $n(n-1)$

est

$$\overline{n(n-1)} = \sum_{n=2}^{N} n(n-1)P_n = N(N-1)p^2$$

Ces valeurs sont bien équivalentes aux moyennes ν et ν^2 trouvées plus haut, quand N augmente indéfiniment, p devenant alors la quantité $\frac{t}{T}$.

L'écart est ici $n - Np$; la fluctuation de son carré est

$$\overline{(n-Np)^2} = \overline{n^2} - 2Np\,\overline{n} + N^2p^2 = Npq.$$

On retrouve ainsi le carré de l'unité d'écart, utilisée dans l'etude de la loi de Gauss; mais ce calcul ne suppose rien sur l'ordre de grandeur de N, et montre que cette relation est vraie pour des lois des ecarts autres que celle de Gauss, elle est, d'ailleurs, genérale.

Pour démontrer sa généralite, nous nous servirons de la loi générale suivante

« Étant données deux variables x et y, independantes, susceptibles de prendre respectivement les valeurs x_1, x_2, ..., x_n, y_1, y_2, ..., y_p, avec des valeurs moyennes nulles, le carré moyen de $x+y$ est égal à la somme des carres moyens de x et de y. »

Soient $\alpha_1, \alpha_2, \ldots, \alpha_n$ les probabilités d'obtenir $x_1, x_2, \ldots, x_n$; on a $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 1$; soient, de même, $\beta_1, \beta_2, \ldots, \beta_p$ les probabilités d'obtenir $y_1, y_2, \ldots, y_p$, avec $\beta_1 + \beta_2 + \ldots + \beta_p = 1$. Par hypothèse,

$$\sum_{i=1}^{n} \alpha_i x_i = 0$$

$$\sum_{k=1}^{p} \beta_k y_k = 0.$$

Enfin, par définition, on a

$$\overline{x^2} = \sum_{i=1}^{n} \alpha_i x_i^2,$$

$$\overline{y^2} = \sum_{k=1}^{p} b_k y_k^2.$$

Ceci posé, on peut considérer toutes les valeurs de $x_i + y_k$ comme distinctes, de sorte que la probabilité d'obtenir $x_i + y_k$ est $\alpha_i \beta_k$. Donc

$$\begin{aligned}
\overline{(x+y)^2} &= \sum_{i=1}^{n} \sum_{k=1}^{p} \alpha_i \beta_k (x_i + y_k)^2 \\
&= \left(\sum_k \beta_k\right)\left(\sum_i \alpha_i x_i^2\right) + \left(\sum_i \alpha_i\right)\left(\sum_k \beta_k y_k^2\right) \\
&\qquad + 2\left(\sum_i \alpha_i x_i\right)\left(\sum_k \beta_k y_k\right) \\
&= \overline{x^2} + \overline{y^2}.
\end{aligned}$$

On voit que la démonstration a le plus grand caractère de généralité, et que la conclusion serait également valable pour des variables continues, ainsi que pour un nombre quelconque de variables.

Appliquons cette loi à deux événements de probabilités p et $q = 1 - p$. Désignons par 1 le cas favorable, par 0 le cas défavorable. La valeur moyenne est p, donc on obtiendra l'écart $x_1 = 1 - p = q$ avec la probabilité p, et l'écart $x_2 = 0 - p = -p$ avec la probabilité q. Le carré moyen de l'écart x susceptible de prendre ces deux valeurs est alors

$$\overline{x^2} = q^2 p + p^2 q = pq.$$

S'il s'agit de N expériences, il suffit de prendre N variables analogues à x, indépendantes entre elles, et d'appliquer la loi que nous avons établie. On en a le droit puisque x a la valeur moyenne zéro, ce qui donne bien

$$\overline{\text{écart}^2} = \overline{(x + x' + \ldots + x^{(N)})^2} = Npq.$$

II. — PROBLÈMES DANS L'ESPACE.

13. Nous allons nous occuper, dans cette deuxième partie, de l'autre mode de généralisation que nous avions indiqué.

Considérons un plan, et, dans ce plan, une aire Σ. Si l'on répartit N points dans Σ, de façon que la probabilité d'un quelconque de ces points, de tomber dans une aire S, soit $\frac{S}{\Sigma}$, on peut définir, comme

pour la droite, une densité superficielle $\sigma = \frac{N}{\Sigma}$, et le nombre moyen de points qui se répartissent dans une aire S est $\nu = \frac{NS}{\Sigma} = \sigma S$.

On peut se proposer les mêmes problèmes que sur la droite, en supposant que N et Σ ont augmenté indéfiniment, tout en restant dans un rapport constant $\frac{N}{\Sigma} = \sigma$. On peut diviser l'aire S en aires égales, comme nous avons fait pour le segment x, les calculs peuvent être conduits suivant les mêmes principes, et les probabilités p_n ne seront pas différentes. En particulier, n et n^2 auront les mêmes valeurs moyennes

Mais les problèmes relatifs à des groupes de points conduisent à des résultats tout de suite distincts et compliqués

Étudions, par exemple, la probabilité pour que le segment joignant deux points A_1, A_2 soit inférieur à une longueur donnée ε Nous suivrons le même mode de raisonnement que sur la droite, en supposant l'aire S suffisamment grande pour que l'influence de sa frontière puisse être négligée

A chaque point A_1 de S, correspond un cercle de centre A_1, et de rayon ε, qui est la portion de S utilisable par A_2. Son aire est $\pi\varepsilon^2$, et le nombre moyen de points, dans ce cercle, est $\pi\sigma\varepsilon^2$ (Pour la même raison que sur la droite, on ne tient pas compte du point A_1 qui se trouve déjà a l'intérieur de ce cercle.) Le nombre moyen de ces cercles étant σS. le nombre moyen des extrémités des segments inférieurs à ε est $\pi\sigma^2\varepsilon^2 S$; leur densité est donc $\pi\sigma^2\varepsilon^2$, et celle des segments, $\frac{1}{2}\pi\sigma^2\varepsilon^2$. Il en résulte que la probabilité cherchée, rapport du nombre moyen des segments inférieurs a ε, au nombre moyen $\frac{\sigma^2 S^2}{2}$ de tous les segments intérieurs à S, est $\frac{\pi\varepsilon^2}{S}$.

On peut encore dire que ces résultats sont vrais rigoureusement, sans négliger la frontière de S, à condition de faire entrer en ligne de compte les segments, inférieurs à ε, qui chevauchent cette frontière.

On arrive au même résultat par le calcul de la partie utile relative à plusieurs points; le raisonnement est le même que pour la droite, et la somme des parties utilisables est, pour n points,

$$\pi\varepsilon^2[1 + 2 + 3 + \ldots + (n-1)] = \frac{n(n-1)}{2}\pi\varepsilon^2.$$

Le nombre probable des segments inférieurs à ε, parmi les $\frac{n(n-1)}{2}$ segments de l'aire S, est donc $\frac{n(n-1)}{2}\frac{\pi\varepsilon^2}{S}$, dont la valeur moyenne est

$$\frac{\pi\nu^2\varepsilon^2}{2S} = \frac{\pi\sigma^2\varepsilon^2}{2}S.$$

C'est le nombre moyen des segments inférieurs à ε que fournirait une suite d'expériences faites avec la même densité superficielle σ. On retrouve enfin la même expression $\frac{\pi\sigma^2\varepsilon^2}{2}$ pour la densité moyenne de ces segments. Nous allons faire une étude plus précise du problème précédent dans quelques cas simples.

14. Cercle. — Ce problème conduit tout de suite à des calculs très compliqués, si l'on veut tenir compte du contour de l'aire S, même si ce contour est aussi simple qu'une circonférence.

« Considérons un cercle de rayon R. Étant donnés deux points A, B, intérieurs à ce cercle, et obéissant à la loi de probabilité continue $\frac{dx\,dy}{S}$, quelle est la probabilité pour que la longueur AB soit inférieure à un nombre donné ε ? »

Soient (x_1, y_1), (x_2, y_2) les coordonnées des deux points A et B, rapportées à deux diamètres rectangulaires du cercle. Les deux inégalités

$$(C_1) \qquad x_1^2 + y_1^2 < R^2,$$

$$(C_2) \qquad x_2^2 + y_2^2 < R^2$$

expriment que A et B sont intérieurs au cercle. Pour que AB soit un segment favorable, il faut, en outre, qu'il satisfasse à l'inégalité

$$(A) \qquad a^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 < \varepsilon^2.$$

Dans ces conditions, la probabilité en question est représentée par le rapport de l'intégrale $\int\int\int\int dx_1\,dy_1\,dx_2\,dy_2$ prise pour les seuls segments favorables, à cette même intégrale prise pour tous les segments possibles. Autrement dit,

$$P = \frac{\int\int\int\int_{C_1, C_2, A} dx_1\,dy_1\,dx_2\,dy_2}{\int\int\int\int_{C_1, C_2} dx_1\,dy_1\,dx_2\,dy_2};$$

l'intégrale du numérateur est prise dans le domaine à quatre dimensions défini par les trois inégalités (C_1), (C_2), (A) et, celle du dénominateur, dans le domaine défini par (C_1) et (C_2).

Cette dernière intégrale a évidemment pour valeur $\pi^2 R^4$; l'autre intégrale peut être calculée analytiquement, mais on peut simplifier beaucoup les calculs grâce à des considérations géométriques.

A étant supposé fixé, AB peut avoir une direction quelconque. L'angle $\theta = (Ox, \overrightarrow{AB})$ peut donc varier de o à 2π, et la longueur $AB = a$ doit être comprise entre o et ε. Considérons alors un intervalle $(a, a + da)$, $(\theta, \theta + d\theta)$, relatif à ces deux variables, et remarquons qu'on peut toujours entourer le point A d'un élément d'aire $dx_1 dy_1$ infiniment petit par rapport à $a\,da\,d\theta$, qui représente l'étendue, indépendante de $dx_1 dy_1$, réservée à B (à moins que cette étendue ne soit extérieure au cercle).

Pour voir si B se trouve dans le cercle, effectuons, sur le cercle donné (C), la translation équipollente au vecteur défini par $(a, \theta + \pi)$ (C) devient le cercle (C'), et B vient en A Pour que B soit intérieur au cercle primitif, il faut et il suffit que A soit dans la région

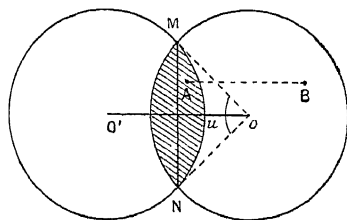



Fig. 11.

commune à ces deux cercles La simplicité du calcul provient de ce que l'aire de cette région est indépendante de la valeur de θ Si, au lieu de a, on prend comme variable l'angle $u = \widehat{MON}$, on a

$$a = 2R\cos\frac{u}{2},$$

et l'aire commune aux deux cercles est égale à

$$R^2(u - \sin u),$$

L'intégrale quadruple cherchée se réduit alors à l'intégrale double

$$\int_0^{2\pi}\int_0^{\varepsilon} R^2(u - \sin u)\, a\, da\, d\theta,$$

et, immédiatement, à l'intégrale simple

$$2\pi R^2 \int_0^{\varepsilon} (u - \sin u) a\, da.$$

L'intervalle de variation de u est $\pi \geqq u \geqq 2 \operatorname{arc\,cos} \frac{\varepsilon}{2R}$. Posons

$$\frac{\varepsilon}{2R} = \sin \frac{\alpha}{2},$$

cet intervalle s'écrit $\pi \geqq u \geqq \pi - \alpha$. Enfin, en exprimant $a\, da$ en fonction de u, on obtient

$$a\, da = - R^2 \sin u\, du,$$

ce qui donne, en définitive,

$$P = \frac{2}{\pi}\int_{\pi-\alpha}^{\pi} (u - \sin u) \sin u\, du$$

Il vient alors

$$P = \frac{2}{\pi}\left[\sin u - u \cos u - \frac{u}{2} + \frac{\sin 2u}{4}\right]_{\pi-\alpha}^{\pi},$$

$$P = \frac{1}{\pi}\left[-2 \sin \alpha + 2\pi - 2(\pi - \alpha)\cos \alpha - \alpha + \frac{\sin 2\alpha}{2}\right]$$

or

$$\cos \alpha = 1 - 2 \sin^2 \frac{\alpha}{2} = \frac{2R^2 - \varepsilon^2}{2R^2},$$

d'où l'on déduit

$$P = \frac{1}{\pi R^2}\left\{2\pi R^2 - (\pi - \alpha)(2R^2 - \varepsilon^2) - \alpha R^2 - \sin \alpha \left(R^2 + \frac{\varepsilon^2}{2}\right)\right\},$$

et enfin

$$P = \frac{\pi\varepsilon^2 + \alpha(R^2 - \varepsilon^2) - \sin \alpha \left(R^2 + \frac{\varepsilon^2}{2}\right)}{\pi R^2}.$$

C'est cette expression de P que nous conserverons, car elle met en évidence une forme générale que nous retrouverons plus loin.

Si R augmente indéfiniment, α tend vers zéro, et la probabilité P est équivalente à la valeur connue $\frac{\pi \varepsilon^2}{\pi R^2}$.

Si α n'est pas négligeable, mais est assez petit pour être assimilé à son sinus, on a

$$P = \frac{\varepsilon^2}{R^2}\left(1 - \frac{3\alpha}{2\pi}\right),$$

probabilité inférieure a celle que l'on obtient en négligeant α; ce résultat est naturel, car dans celle-ci on fait entrer en ligne de compte les cercles $\pi\varepsilon^2$ qui coupent la circonférence du cercle (C)

Dans le cas particulier où ε atteint son maximum $2R$, α est egal à π, et l'on vérifie bien que $P = 1$; enfin, si ε est egal à R, α a pour valeur $\frac{\pi}{3}$, ce qui donne $P = 1 - \frac{3\sqrt{3}}{4\pi}$.

15 **Sphère.** — La méthode précédente s'applique, sans modification, a la sphère, il se produit, cependant, certaines simplifications dans les calculs.

On rapportera les points A et B à trois diamètres rectangulaires de la sphère; soient (a, θ, φ) les cordonnées polaires du vecteur $\overrightarrow{AB}$, les intervalles de variation de ces variables sont $0 \leqq a \leqq \varepsilon$, $0 \leqq \theta \leqq 2\pi$, $0 \leqq \varphi \leqq \pi$. Si R est le rayon de la sphère, le volume commun a cette sphère et à celle qu'on en déduit par la translation $\overrightarrow{BA}$ a pour mesure

$$\frac{\pi}{3}\left(4R^3 - 3aR^2 + \frac{a^3}{4}\right),$$

donc la probabilité P est ici

$$P = \frac{1}{\frac{16}{9}\pi^2 R^6}\int_0^{2\pi}\int_0^{\pi}\int_0^{\varepsilon}\frac{\pi}{3}\left(4R^3 - 3aR^2 + \frac{a^3}{4}\right)a^2 \sin\varphi \, da \, d\theta \, d\varphi,$$

$$P = \frac{3}{4R^6}\int_0^{\varepsilon}\left(4R^3 - 3R^2 a + \frac{a^3}{4}\right)a^2 \, da,$$

ou enfin

$$P = \left(\frac{\varepsilon}{R}\right)^3 - \left(\frac{3}{4}\frac{\varepsilon^2}{R^2}\right)^2 + \frac{1}{32}\frac{\varepsilon^6}{R^6},$$

qui ne contient aucune quantité transcendante.

16. **Triangle.** — Ce qui fait la simplicité du problème relatif au cercle, c'est l'absence de directions privilégiées. Il n'en est plus de même pour un polygone.

Considérons, par exemple, un triangle ABC, de côtés a, b, c et

d'angles au sommet A, B, C. Désignons par L son périmètre, et déterminons, en fonction de ces éléments, la probabilité pour qu'un segment M_1M_2, intérieur au triangle, ait une longueur inférieure à z.

Soient $d\omega_1$, $d\omega_2$ les aires infinitésimales associées aux deux points M_1, M_2. Comme pour le cercle, nous avons

$$P = \frac{\int\int_{(1),(2)} d\omega_1\, d\omega_2}{\int\int_{(1)} d\omega_1\, d\omega_2},$$

(1) désignant les inégalités qui expriment que M_1 et M_2 sont intérieurs au triangle, et (2) l'inégalité qui exprime que M_1M_2 est au plus égal à z. Le dénominateur est donc égal au carré S^2 de l'aire du triangle.

Pour le calcul du numérateur, nous emploierons encore la méthode géométrique. Remarquons tout d'abord que l'on peut toujours mener une direction parallèle à M_1M_2, par l'un des sommets du triangle, de façon que cette direction soit intérieure à l'angle de ce sommet. On peut partager les segments M_1M_2 en trois catégories, suivant l'angle à l'intérieur duquel se trouvera la direction de ce segment. L'intégrale $\int\int_{(1)\ (2)} d\omega_1 d\omega_2$ est alors la somme des intégrales relatives à ces trois catégories.

Calculons-la pour l'angle A. Désignons par ρ la longueur M_1M_2, et

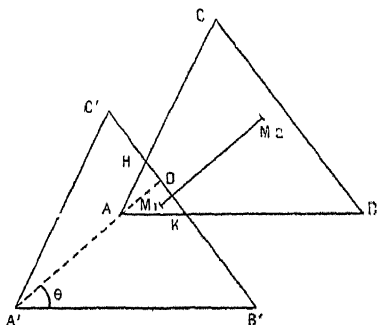



Fig. 12.

par θ l'angle $(\overrightarrow{AB}, \overrightarrow{M_1M_2})$. θ est compris entre o et A, ou entre π et $A+\pi$. L'intégrale cherchée est évidemment le double de celle que l'on obtient en limitant la variation de θ à l'intervalle (o, A).

Imprimons au triangle ABC la translation définie par le vecteur $\overrightarrow{M_2 M_1}$, et soit $A'B'C'$ le triangle égal ainsi obtenu. M_1 doit se trouver dans l'aire commune à ces deux triangles, si cette aire existe.

Si $A'A$ coupe $B'C'$ en D, il faut, pour que cette aire existe, que $A'A$ soit inférieur à $A'D$, longueur qui est une fonction connue de θ. Donc ρ doit varier entre 0 et ε si ε est inférieur à $A'D$, et entre 0 et $A'D$, dans le cas contraire Les calculs seront donc complexes, sauf si ε est toujours inférieur à $A'D$, quel que soit l'angle θ Pour que la même simplification se produise dans chacun des trois angles au sommet, on est conduit à supposer que ε ne dépasse pas la plus petite hauteur du triangle ABC C'est ce que nous ferons.

Ceci posé, on a

$$d\omega_2 = \rho\, d\rho\, d\theta,$$

et $\int d\omega_1 =$ aire AKH est une fonction de ρ et θ que nous allons déterminer. La probabilité relative a l'angle A est alors

$$P_A = \frac{2}{S^2}\int_0^A \int_0^\varepsilon (\text{aire AKH})\rho\, d\rho\, d\theta$$

Pour calculer l'aire AKH, écrivons qu elle est homothétique au triangle $A'B'C'$; il vient

$$\frac{\text{aire AKH}}{S} = \left(\frac{AD}{A'D}\right)^2 = \left(1 - \frac{\rho}{A'D}\right)^2;$$

or

$$A'D = \frac{c \sin B}{\sin(B+\theta)},$$

d'où résulte

$$\text{aire AKH} = S\left[1 - \frac{\rho \sin(B+\theta)}{c \sin B}\right]^2.$$

On peut donc écrire

$$P_A S^2 = 2S\int_0^A \int_0^\varepsilon \rho\left(1 - \frac{2\rho}{c}\,\frac{\sin(B+\theta)}{\sin B} + \frac{\rho^2}{c^2}\,\frac{\sin^2(B+\theta)}{\sin^2 B}\right) d\rho\, d\theta.$$

Les intégrations par rapport à ρ, puis par rapport à θ, donnent successivement

$$P_A . S^2 = S\int_0^A \left[\varepsilon^2 - \frac{4}{3}\,\frac{\varepsilon^3 \sin(B+\theta)}{c \sin B} + \frac{1}{2}\,\frac{\varepsilon^4 \sin^2(B+\theta)}{c^2 \sin^2 B}\right] d\theta$$

$$= S\varepsilon^2 A - \frac{4}{3}\varepsilon^3 S\,\frac{\cos B + \cos C}{c \sin B} + \frac{1}{8}\varepsilon^4 S\,\frac{2A + \sin 2B + \sin 2C}{c^2 \sin^2 B}.$$

Or

$$S = \frac{1}{2} ac \sin B$$

donne

$$S \frac{\cos B + \cos C}{c \sin B} = \frac{a}{2} (\cos B + \cos C)$$

D'autre part, on peut ne conserver dans l'expression du dernier terme que les angles A, B, C, grâce aux trois identités

$$\frac{2AS}{c^2 \sin^2 B} = \frac{Aa}{c \sin B} = \frac{A \sin A}{\sin B \sin C} = A(\cot B + \cot C),$$

$$\frac{S \sin 2B}{c^2 \sin^2 B} = \frac{a \sin 2B}{2c \sin B} = \frac{a \cos B}{c} = \frac{\sin A \cos B}{\sin C},$$

$$\frac{S \sin 2C}{c^2 \sin^2 B} = \frac{a \sin 2C}{2c \sin B} = \frac{a \sin 2C}{2b \sin C} = \frac{\sin A \cos C}{\sin B}.$$

On peut donc, en définitive, mettre $P_A S^2$ sous la forme

$$P_A S^2 = S\varepsilon^2 A - \frac{2}{3}\varepsilon^3 a(\cos B + \cos C)$$
$$+ \frac{\varepsilon^4}{8}\left[\frac{\sin A \cos B}{\sin C} + \frac{\sin A \cos C}{\sin B} + A(\cot B + \cot C)\right].$$

Si nous ajoutons maintenant les expressions relatives aux deux autres angles B et C, qui se déduisent de celle-ci par permutation circulaire, et si nous groupons les termes suivant les puissances de ε, nous obtenons enfin :

$$\begin{array}{rl|ll|lll}
PS^2 = & S\varepsilon^2 & A & -\frac{2}{3}\varepsilon^3 & a\cos B + a\cos C & & \\
& & + B & & b\cos A & & + b\cos C \\
& & + C & & & + c\cos A & + c\cos B
\end{array}$$

$$\begin{array}{r|llll}
+\frac{\varepsilon^4}{8} & A(\cot B + \cot C) + \frac{\sin A \cos B}{\sin C} + \frac{\sin A \cos C}{\sin B} & & & \\
& + B(\cot C + \cot A) + \frac{\sin B \cos A}{\sin C} & & & + \frac{\sin B \cos C}{\sin A} \\
& + C(\cot A + \cot B) & & + \frac{\sin C \cos A}{\sin B} & + \frac{\sin C \cos B}{\sin A}.
\end{array}$$

Additionnons tous ces termes, colonne par colonne, en remarquant que

$$a \cos B + b \cos A = c$$

et que

$$\frac{\sin A \cos B}{\sin C} + \frac{\sin B \cos A}{\sin C} = 1$$

ce qui nous donne finalement

$$P(\varepsilon) = \frac{1}{S^2}\Big\{\pi S \varepsilon^2 - \frac{2}{3} L \varepsilon^3 + \frac{\varepsilon^4}{8} \times [(\pi - A)\cot A + (\pi - B)\cot B + (\pi - C)\cot C + 3]\Big\}.$$

Cette probabilité contient trois termes, en ε^2, ε^3, et ε^4.

Le premier ne dépend que de l'aire du triangle.
Le deuxième ne dépend que de son périmètre.
Le troisième ne dépend que de ses angles.

Nous allons voir que cette constatation n'est pas particulière au triangle, mais s'applique, d'une manière générale, à tous les polygones convexes.

17. **Carré** — Avant d'examiner le cas général, déterminons encore l'expression de cette probabilité, pour un carré. Considérons un carré de côté a, son aire S a pour mesure a^2, et son périmètre est $L = 4a$. Par contre, le terme en ε^4 devra avoir un coefficient numérique puisque les angles sont ici déterminés.

Cette probabilité se calcule toujours par le même artifice, mais une grande simplification de calcul résulte de ce qu'on peut toujours mener, par l'un des quatre sommets A, B, C, D du carré, un vecteur équipollent à $\overrightarrow{M_1 M_2}$, et intérieur au carré.

L'aire commune à ABCD et au carré A′ B′ C′ D′ déduit de ABCD par la translation $\overline{M_2 M_1}$ a pour mesure

$$(a - \rho\cos\theta)(a - \rho\sin\theta).$$

Si l'on suppose toujours que la limite supérieure de ρ est ε, indépendamment de la valeur de l'angle θ, ce qui revient à supposer que ε est inférieur ou égal au côté a, on a

$$PS^2 = 4\int_0^{\frac{\pi}{2}}\int_0^{\varepsilon}(a - \rho\cos\theta)(a - \rho\sin\theta)\,\rho\, d\rho\, d\theta.$$

L'intégration par rapport à ρ, puis par rapport à θ, donne

$$\begin{aligned} PS^2 &= \pi a^2 \varepsilon^2 - \frac{4a\varepsilon^3}{3}\int_0^{\frac{\pi}{2}} (\sin\theta + \cos\theta)\, d\theta + \varepsilon^4 \int_0^{\frac{\pi}{2}} \sin\theta \cos\theta\, d\theta \\ &= \pi \varepsilon^2 S - \frac{8a\varepsilon^3}{3} + 2\frac{\varepsilon^4}{4}, \end{aligned}$$

ce qui s'écrit encore

$$P = \frac{1}{S^2}\left\{ \pi S \varepsilon^2 - \frac{2}{3} L \varepsilon^3 + \frac{\varepsilon^4}{8} \cdot 4 \right\}.$$

Cette expression de la probabilité est bien de la forme annoncée

18. Discussion Polygone convexe. — Nous nous proposons, dans ce paragraphe, d'étudier la signification de chacun des trois termes qui entrent dans l'expression de la probabilité trouvée pour le triangle, et d'en déduire l'expression de cette probabilité pour un polygone convexe quelconque.

Si ε est infiniment petit par rapport au triangle, la probabilité P est équivalente à $\frac{\pi\varepsilon^2}{S}$. C'est la probabilité que l'on obtient, pour une aire quelconque de mesure S, lorsqu'on ne tient aucun compte de la frontière.

Le deuxième terme $-\frac{2}{3}\frac{L\varepsilon^3}{S^2}$ ne fait intervenir que l'aire et le périmètre; sa valeur absolue mesure, en effet, les portions d'aire, extérieures au triangle, de tous les cercles $\pi\varepsilon^2$ qui sont traversés par la frontière, de longueur L, et qui sont comptés, en trop, dans le premier terme. Pour le vérifier, calculons cette aire, en remplaçant la frontière du triangle par un segment de droite de longueur L, sans tenir compte, dans ce calcul, de l'influence des extrémités. La correction due aux angles du triangle sera évaluée séparément.

Soit alors un point M, situé au-dessus de la droite de longueur L, à une distance de cette droite inférieure à ε. Quand, dans le calcul du premier terme, on associe à ce point l'aire $\pi\varepsilon^2$ du cercle de rayon ε dont il est le centre, on ajoute, en trop, l'aire hachurée située au-dessous de la droite.

Désignons par 2α l'angle sous lequel on voit, de M, le segment de frontière limité par le cercle, et soient (x, y) les coordonnées cartésiennes de M rapportées à la droite, et à un axe orthogonal, choisi

arbitrairement. L'aire hachurée a pour mesure

$$\varepsilon^2\left(\alpha - \frac{\sin 2\alpha}{2}\right).$$

On corrigera l'erreur que l'on faisait en négligeant l'influence de la frontière (due à sa longueur seule), en retranchant, du premier terme, l'aire totale

$$\int\int \varepsilon^2(\alpha - \sin\alpha\cos\alpha)\,dx\,dy$$

x est indépendant de y, et l'amplitude de sa variation est égale à L, puisqu'on ne tient pas compte des extrémités, d'autre part, en

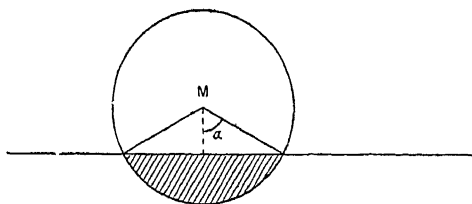



Fig 13.

prenant pour variable l'angle α, variant de 0 à $\frac{\pi}{2}$, on a $y = \varepsilon\cos\alpha$, $dy = -\varepsilon\sin\alpha\,d\alpha$, et, par suite, pour expression de l'intégrale,

$$L\varepsilon^3\int_0^{\frac{\pi}{2}}(\alpha\sin\alpha - \sin^2\alpha\cos\alpha)\,d\alpha = L\varepsilon^3\left(-\alpha\cos\alpha + \sin\alpha - \frac{\sin^3\alpha}{3}\right)_0^{\frac{\pi}{2}} = \frac{2}{3}L\varepsilon^3$$

Nous obtenons ainsi un premier terme correctif $-\frac{2}{3}L\varepsilon^3$, qui nous fournit la signification du deuxième terme de $P(\varepsilon)$.

Examinons maintenant quelle sera l'influence des angles, dont nous n'avons tenu aucun compte jusqu'ici. Considérons l'un des angles du triangle, A par exemple.

Rappelons-nous que nous avons déterminé le terme $\pi S\varepsilon^2$, qui représente la somme des aires des cercles de rayon ε, dont le centre est dans le triangle ABC, et que nous venons d'en retrancher la mesure des portions, extérieures au triangle, des cercles de rayon ε dont le centre se projette sur un côté.

Lorsqu'on retranche ce terme correctif $\frac{2}{3}L\varepsilon^3$, on retranche deux fois les aires telles que AMN, relatives à un point P intérieur à

l'angle A, on retranche également, en trop, les aires telles que MQN relatives aux centres P, extérieurs au triangle, et dont l'une des projections est intérieure à un côté.

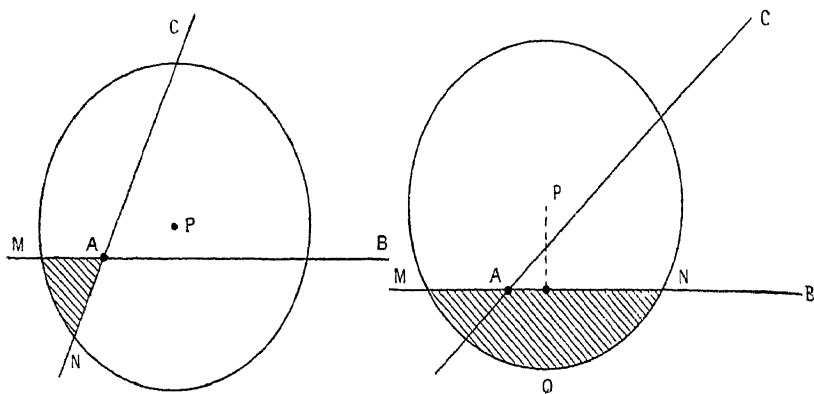



Fig. 14.

Il faut donc corriger le terme correctif $-\frac{2}{3}L\varepsilon^3$, par un terme additif. Il est clair que la somme des aires de cette nature est de l'ordre de ε^4, car l'aire limitée par un cercle est de l'ordre de ε^2, et l'intégrale $\int\int dx\,dy$ étendue à l'ensemble de ces points P est également de l'ordre de ε^2. Le terme correctif relatif à l'angle A est donc de la forme $\frac{\varepsilon^4}{8}\varphi(A)$, la fonction φ étant bien déterminée, et sa valeur ne dépendant que de l'angle A et non du polygone dont cet angle fait partie.

On peut la déterminer par un calcul direct; mais ce calcul est très compliqué, et il est beaucoup plus simple d'utiliser les calculs effectués plus haut pour les cas particuliers du triangle et du carré.

Pour le triangle, le terme correctif total est de la forme

$$\frac{\varepsilon^4}{8}[\varphi(A)+\varphi(B)+\varphi(C)],$$

où $A+B+C=\pi$, et nous connaissons sa valeur

$$\frac{\varepsilon^4}{8}[(\pi-A)\cot A+(\pi-B)\cot B+(\pi-C)\cot C+3].$$

Posons alors

$$f(A)=(\pi-A)\cot A+1;$$

Ceci nous permet de dire que la relation $A+B+C=\pi$ entraîne l'identité

$$\varphi(A)+\varphi(B)+\varphi(C)=f(A)+f(B)+f(C)$$

ou

$$[\varphi(A)-f(A)]+[\varphi(B)-f(B)]+[\varphi(C)-f(C)]=0;$$

on peut dire encore, en posant $A=\frac{\pi}{3}+A'$, et $\varphi(A)-f(A)=\psi(A')$, que $A'+B'+C'=0$ entraîne

$$\psi(A')+\psi(B')+\psi(C')=0$$

Cette condition revient à définir $\psi(x)$ par l'équation fonctionnelle

$$\psi(x+y)=-\psi(-x)-\psi(-y),$$

et cette fonction $\psi(x)$ est continue, puisque $\varphi(x)$ et $f(x)$ le sont. Il résulte immédiatement de cette équation que $\psi(x)$ est de la forme kx, k étant une constante, ce qui donne

$$\varphi(A)=f(A)+k\left(A-\frac{\pi}{3}\right).$$

Nous allons montrer que la constante k est nulle. Pour le voir, il suffit d'identifier la somme

$$\frac{\varepsilon^4}{8}\sum\varphi(A)=\frac{\varepsilon^4}{8}\left[\sum f(A)+k\sum\left(A-\frac{\pi}{3}\right)\right],$$

étendue aux angles d'un polygone pour lequel le coefficient de k soit différent de zéro, avec le troisième terme de la probabilité calculée directement pour ce polygone. Il suffit d'utiliser le calcul relatif au carré; dans ces conditions, $A=B=C=D=\frac{\pi}{2}$, et $\frac{\varepsilon^4}{8}\Sigma f(A)=4\frac{\varepsilon^4}{8}$ est identique au troisième terme en question.

L'identification conduit donc à l'égalité zéro de

$$k\sum\left(A-\frac{\pi}{3}\right)=k\frac{\pi}{6},$$

c'est-à-dire à $k=0$. C. Q. F. D.

En résumé, la formule déterminée pour le triangle est de forme absolument générale. Pour un polygone convexe d'un nombre quelconque de côtés, la probabilité pour qu'un segment M_1M_2 soit inférieur à ε est donc

$$P(\varepsilon)=\frac{1}{S^2}\left\{\pi S\varepsilon^2-\frac{3}{2}L\varepsilon^3+\frac{\varepsilon^4}{8}\sum[(\pi-A)\cot A+1]\right\},$$

où S et L sont l'aire et le périmètre de ce polygone, et où la dernière somme est étendue à tous les angles de ce polygone.

Il ne faut pas oublier, cependant, que ε doit être suffisamment petit par rapport aux côtés du polygone pour qu'une longueur ε, intérieure au polygone, et ayant un sommet quelconque pour origine, ne puisse rencontrer le contour de ce polygone en un autre point.

19. Remarque. — Il est curieux de déduire le premier terme correctif $\frac{2}{3}\frac{L\varepsilon^3}{S^2}$ du terme principal $\pi S\varepsilon^2$, en faisant appel au problème de l'aiguille de Buffon.

Considérons n droites parallèles, de longueur commune l, régulièrement espacées, l'écart commun étant d, l est assez grand pour que les extrémités de ces droites puissent être négligées, et n assez grand pour que $n \pm 1$ puisse être considéré comme équivalent à n.

Supposons maintenant que deux points soient placés au hasard dans ce rectangle réglé. On sait que la probabilité pour que le segment de ces deux points ne surpasse pas une longueur donnée ρ est $P(\rho) = \frac{\pi\rho^2}{S}$; donc la probabilité pour qu'un pareil segment soit compris entre ρ et $\rho + d\rho$ est $dP(\rho) = \frac{2\pi\rho\, d\rho}{S}$.

L'espérance mathématique d'un joueur qui perdrait l'unité pour chaque point de rencontre d'un segment inférieur à ε avec l'une des droites est

$$-\int_0^{\varepsilon} \frac{2\rho}{\pi d} \frac{2\pi\rho\, d\rho}{S} = -\frac{4}{3}\frac{\varepsilon^3}{dS}.$$

Or la longueur totale de la frontière constituée par ces n droites est équivalente à $L = 2nl$, car chaque droite, autre que les extrêmes, est frontière de deux rectangles; d'autre part S est égal à nld, ce qui donne, pour valeur de cette probabilité,

$$-\frac{2}{3}\frac{L\varepsilon^3}{nl\, dS} = -\frac{2}{3}\frac{L\varepsilon^3}{S^2}.$$

20. Espace à trois dimensions. — Nous avons déjà étudié le problème général qui nous occupe, dans l'espace à trois dimensions. Mais la simplicité que nous avons rencontrée à propos de la sphère est tout à fait spéciale à ce cas particulier, et l'on rencontre, au con-

traire, de grandes difficultés, dès qu'il s'agit de polyèdres. Pour concevoir quelle peut être la complexité de cette question, il suffit d'examiner ce qui se produit pour un tétraèdre

En effet, les quatre trièdres d'un tétraèdre ne remplissent pas tout l'espace dirigé On le voit schématiquement en menant par un point O des plans parallèles aux quatre faces, et en représentant les traces de ces plans sur une sphère de centre O

Sur la figure, le contour apparent de la sphère est supposé représenter la trace du plan diamétral parallèle à l'une des faces, BCD par exemple, et les grands cercles $\beta\gamma$, $\gamma\delta$, $\delta\beta$, les traces des plans parallèles aux trois autres faces ABC, ACD, ADB

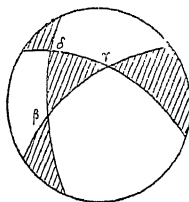



Fig. 15

On voit immédiatement que la parallèle à toute direction intérieure à l'un des quatre angles au sommet du tétraèdre, menée par O, perce la sphère dans l'une des régions hachurées; et ces régions ne recouvrent pas entièrement la demi-sphère

Cette difficulté générale ne se rencontre pas, exceptionnellement, avec le prisme; les angles au sommet d'un tel polyèdre contiennent en effet toutes les directions de l'espace. Nous sommes donc conduits à penser que la question doit être encore relativement simple dans ce cas particulier C'est ce qui a lieu, et nous allons, à titre d'exercice, effectuer le calcul relatif au cube

Nous emploierons le même raisonnement que pour le carré, et nous supposerons que la longueur ε ne dépasse pas le côté a du cube. La probabilité en question est huit fois celle que fournissent les segments M_1M_2 équipollents aux vecteurs intérieurs à un seul des angles au sommet. En rapportant ce vecteur aux arêtes de cet angle, les intervalles de variation de ses coordonnées polaires (ρ, θ, φ) sont $0 \leqq \rho \leqq \varepsilon$, $0 \leqq \theta \leqq \frac{\pi}{2}$, $0 \leqq \varphi \leqq \frac{\pi}{2}$ D'autre part, $\int\int\int d\omega_1$ étendue au volume commun au cube donné et au cube qu'on en déduit par la

translation $\overrightarrow{M_2 M_1}$ est

$$(a - \rho\cos\theta)\,(a - \rho\sin\theta\cos\varphi)\,(a - \rho\sin\theta\sin\varphi),$$

et l'élément de volume $d\omega_2$ associé à l'extrémité M_2 a pour expression $\rho\sin\theta\,d\rho\,d\theta\,d\varphi$.

On a donc

$$P(\varepsilon) = \frac{8}{V^2}\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}}\int_0^{\varepsilon} (a - \rho\cos\theta)\,(a - \rho\sin\theta\cos\varphi) \times (a - \rho\sin\theta\sin\varphi)\,\rho^2\sin\theta\,d\theta\,d\varphi\,d\rho.$$

L'intégration par rapport à ρ donne

$$\begin{aligned} P(\varepsilon) = \frac{8}{V^2}\int_0^{\frac{\pi}{2}}\int_0^{\frac{\pi}{2}} \Big[& \frac{a^3\varepsilon^3}{3} - \frac{a^2\varepsilon^4}{4}(\cos\theta + \sin\theta\cos\varphi + \sin\theta\sin\varphi) \\ & + \frac{a\varepsilon^5}{5}(\sin\theta\cos\theta\cos\varphi \\ & \quad + \sin\theta\cos\theta\sin\varphi + \sin^2\theta\sin\varphi\cos\varphi) \\ & - \frac{\varepsilon^6}{6}\cos\theta\sin^2\theta\sin\varphi\cos\varphi\Big]\sin\theta\,d\theta\,d\varphi \end{aligned}$$

et enfin, en effectuant les deux dernières intégrations, il vient

$$P(\varepsilon) = \frac{1}{V^2}\left\{\frac{4}{3}\pi\varepsilon^3 V - \frac{\pi}{4}S\varepsilon^4 + L\frac{\varepsilon^5}{5} - \frac{\varepsilon^6}{6}\right\}.$$

Remarquons que le premier terme est l'analogue de celui trouvé pour les polygones du plan, l'aire $\pi\varepsilon^2$ étant simplement remplacée par le volume $\frac{4}{3}\pi\varepsilon^3$ de la sphère de rayon ε. Ce premier terme représente la probabilité que l'on obtiendrait en négligeant la frontière du cube.

Le terme suivant est le terme correctif dû à la surface de ce cube, et se calculerait directement par le même procédé que dans le plan. Le troisième terme corrige l'influence des arêtes, que le deuxième terme exagérait, et, enfin, le dernier terme $\frac{\varepsilon^6}{6}$ compense la correction du troisième terme, que l'existence des sommets rend trop forte.

Nous bornerons là ces considérations élémentaires, qui suffisent à faire comprendre la nature des raisonnements employés dans les problèmes de ce genre, et dont la difficulté de résolution est surtout analytique.

CHAPITRE V.

JEU DE PILE OU FACE

1. Le jeu de pile ou face, dont le principe est si simple, possède un très grand caractère de généralité, et conduit, lorsqu'on l'étudie en détail, aux Mathématiques les plus élevées. Nous nous proposons, dans ce Chapitre, d'en entreprendre l'étude systématique, en utilisant un schema géométrique, intéressant et commode.

Nous conviendrons de représenter par le chiffre 0 un coup perdu et par 1 un coup gagné. Ainsi, une partie, formée d'une succession de coups de pile ou face sera représentée par une succession de chiffres 0 et 1, comme un nombre en numeration binaire.

Par exemple, la partie 01100101 sera représentée par le nombre binaire 0,01100101.

L'etude du jeu de pile ou face et celle de ces nombres binaires apparaissent ainsi comme identiques, mais la numération binaire n'est pas d'un grand secours, bien qu'elle soit théoriquement plus simple que la numération décimale, elle est pratiquement plus compliquée, par suite de l'habitude qu'a prise notre esprit de raisonner avec cette dernière.

Il est plus simple, pour entreprendre l'étude du jeu de pile ou face, d'adopter une représentation geométrique, que nous allons, maintenant, exposer en détail.

Considérons un quadrillage dont les points sont rapportés à deux axes rectangulaires Ox, Oy, appartenant à ce quadrillage; les coordonnées (a, b) de ces points sont alors des nombres entiers. Nous n'utiliserons, d'ailleurs, en général, que les points a coordonnées positives.

Un coup de pile ou face peut être représenté par un segment unité de ce quadrillage, dirigé toujours dans le sens des coordonnées croissantes : par exemple, 1 correspondra à un vecteur unité paral-

lèle à Ox, et σ à un vecteur unité parallèle à Oy. Dans ces conditions, une partie quelconque sera représentée par un chemin, brisé

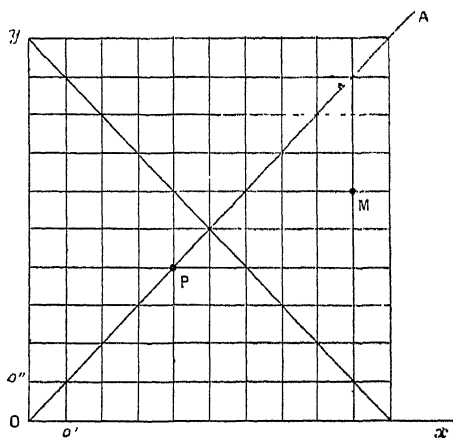



Fig. 16

en général, d'origine O, terminé en un certain point M, et tel qu'on se déplace toujours dans le sens positif des axes quand on suit ce chemin de O vers M.

Si (a, b) sont les coordonnées de l'extrémité M, $a+b$ est le nombre de coups de la partie, a le nombre des coups gagnés, et b le nombre des coups perdus. Toutes les parties, formées d'un même nombre N de coups, ont leurs extrémités sur une diagonale, d'équation $x+y=N$, parallèle à la deuxième bissectrice des axes.

Le nombre total des chemins possibles entre O et M, qui n'est autre que le nombre de parties possibles, qui se terminent avec a coups gagnés et b coups perdus, est

$$N = \frac{(a+b)!}{a!\,b!}.$$

Les parties, formées de b coups gagnés et a coups perdus, sont représentées par les chemins symétriques des précédents par rapport à la première bissectrice OA du quadrillage. OA est le lieu des points M qui correspondent aux parties nulles.

2. Un point M étant fixé, hors de OA, on peut chercher le nombre de chemins qui vont de O vers M, sans rencontrer OA.

Cette question est, sous une autre forme, le problème du scrutin, résolu par M. Désiré André, et que nous avons développé dans le deuxième Chapitre. Si M est au-dessous de OA, $(a > b)$, nous savons que ce nombre est égal à $N \frac{a-b}{a+b}$.

Le raisonnement, qui nous avait conduits à la probabilité $\frac{a-b}{a+b}$, devient intuitif avec ce schéma. A tout chemin, qui coupe OA en un certain point P, correspond le chemin obtenu en remplaçant la portion qui joint O à P, par sa figure symétrique par rapport à OA. Si le premier chemin débute par OO', le deuxième debute par OO''. Or, si M est au-dessous de OA, tout chemin qui commence par OO' coupe nécessairement OA avant d'atteindre M. Leur nombre est le nombre de chemins joignant O'' à M, c'est-à-dire

$$\frac{(a+b-1)!}{a!(b-1)!} = N\frac{b}{a+b}$$

Le nombre des chemins qui vont de O vers M en rencontrant OA est le double, et le nombre de ceux qui ne rencontrent pas OA est alors

$$N\left(1-\frac{2b}{a+b}\right) = N\frac{a-b}{a+b}.$$

3. Étudions, en particulier, les parties nulles. Soit $2a$ le nombre de coups d'une telle partie. Son extrémité M, de coordonnées (a, a), est sur OA. Le nombre de chemins possibles est

$$N_{2a} = \frac{(2a)!}{(a!)^2}.$$

Dans ce cas particulier, il n'y a évidemment aucun chemin qui aille de O vers M sans rencontrer OA; mais on peut rechercher le nombre de ceux qui ne rencontrent pas cette bissectrice entre leurs deux extrémités O et M. Autrement dit, si deux joueurs conviennent de s'arrêter de jouer dès que se produit l'égalité, combien de parties distinctes se terminent au $(2a)^e$ coup?

Le chemin doit rester d'un même côté de OA. Si c'est au-dessous, le $(2a-1)^e$ coup conduit nécessairement au point M', de coordonnées $(a, a-1)$.

Les chemins cherchés sont donc ceux qui aboutissent en M' sans

rencontrer OA, leur nombre est

$$\frac{(2a-1)!}{a!(a-1)!}\frac{1}{2a-1}=\frac{N_{2a}}{2(2a-1)}$$

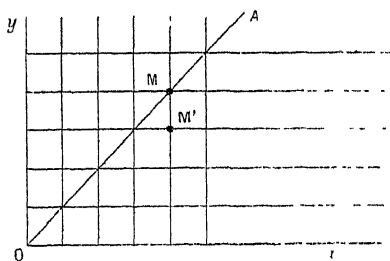



Fig. 17.

Le nombre de tous les chemins qui ne rencontrent pas OA entre O et M est le double, c'est-à-dire

$$\frac{N_{2a}}{2a-1}.$$

La probabilité pour qu'une partie soit nulle à la suite de $2a$ coups est

$$P_{2a}=\frac{N_{2a}}{2^{2a}}=\frac{(2a)!}{2^{2a}(a!)^2},$$

ou encore

$$P_{2a}=\frac{1.3.5\ldots(2a-1)}{2.4.6\ldots2a}.$$

La probabilité pour qu'une partie dure effectivement $2a$ coups, si l'on a décidé de s'arrêter dès la première égalité, est alors

$$P'_{2a}=\frac{1}{2^{2a}}\frac{N_{2a}}{2a-1}=\frac{P_{2a}}{2a-1},$$

ou

$$P'_{2a}=\frac{1.1.3.5\ldots(2a-3)}{2.4.6.8\ldots2a}.$$

Si le nombre de coups est très grand, on a, asymptotiquement,

$$P_{2a}=\frac{(2a)^{2a}e^{-2a}\sqrt{4\pi a}}{a^{2a}e^{-2a}.2\pi a.2^{2a}}=\frac{1}{\sqrt{\pi a}}\quad [1],$$

---

[1] On retrouve ce résultat avec la loi normale des écarts. P_{2a} est en effet la

et, par suite,

$$P'_{2a} = \frac{1}{2a\sqrt{\pi a}}.$$

Ces deux probabilités tendent vers zéro lorsque a augmente indéfiniment, ce qui était bien évident *a priori*

4. **Problème.** — Comme application immédiate des résultats que nous venons d'obtenir, proposons-nous de déterminer la probabilité pour qu'une partie, que l'on arrête dès l'égalité, se termine après $2a$ coups au plus.

C'est évidemment la somme des probabilités

$$P'_2 + P'_4 + P'_6 + \dots + P'_{2a}$$

Celles-ci ont respectivement pour valeurs

$$P'_2 = \frac{1}{2},$$

$$P'_4 = \frac{1}{2}\cdot\frac{1}{4},$$

$$P'_6 = \frac{1}{2}\cdot\frac{1}{4}\cdot\frac{3}{6},$$

$$\dots\dots,$$

$$P'_{2a} = \frac{1}{2}\cdot\frac{1}{4}\cdot\frac{3}{6}\cdots\frac{2a-3}{2a}.$$

Au lieu d'effectuer directement cette somme, il est plus simple, suivant un procédé qui nous est familier, de déterminer la probabilité négative, à savoir la probabilité pour que la partie ne se termine pas avant $2a+2$ coups.

Considérons alors une telle partie. Elle ne doit pas se terminer au bout de 2 coups, ce qui donne la probabilité

$$Q'_2 = 1 - P'_2,$$

ensuite, elle ne doit pas se terminer au quatrième coup, événement

probabilité d'avoir un écart x nul, dans un jeu où $p = q = \frac{1}{2}$, avec $n = 2a$ coups. On a bien

$$P_{2a} = \frac{1}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{\pi a}}.$$

dont la probabilité est

$$Q'_4 = Q'_2 - P'_4,$$

et ainsi de suite jusqu'à

$$Q'_{2a} = Q'_{2a-2} - P'_{2a}.$$

La somme de toutes ces relations fournit l'identité

$$Q'_{2a} = 1 - (P'_2 + P'_4 + \ldots + P'_{2a}).$$

Nous sommes donc ramenés à calculer Q'_{2a}, et l'expression de cette probabilité se met immédiatement sous une forme récurrente simple. On peut écrire, en effet, de proche en proche,

$$Q'_2 = 1 - \frac{1}{2} = \frac{1}{2},$$

$$Q'_4 = \frac{1}{2} - \frac{1}{2}\,\frac{1}{4} = \frac{1}{2}\cdot\frac{3}{4},$$

$$Q'_6 = \frac{1}{2}\cdot\frac{3}{4} - \frac{1}{2}\cdot\frac{1}{4}\cdot\frac{3}{6} = \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6},$$

et, d'une manière générale,

$$Q'_{2a} = \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdots\frac{2a-1}{2a}.$$

On a donc

$$Q'_{2a} = P_{2a},$$

et la probabilité cherchée est $1 - P_{2a}$.

Autrement dit, on a la même probabilité de n'obtenir aucune égalité avant le $(2a+1)^e$ coup que d'en avoir une au $(2a)^e$ coup.

Si a augmente indéfiniment, P_{2a} tend vers zéro, et $P'_2 + P'_4 + \ldots + P'_{2a}$ tend vers 1. Il résulte de cette remarque que P'_∞, probabilité pour que l'égalité ne se produise jamais, est nulle [1].

5. **Durée moyenne.** — Si nous convenons toujours d'arrêter le jeu dès que se produit l'égalité, quelle sera la durée moyenne de la partie?

---

(1) L'énoncé rigoureux de cette propriété serait : la probabilité pour que l'égalité ne se produise qu'après le $(2a)^e$ coup est infiniment petite lorsque $2a$ augmente indéfiniment.

Nous avons désigné par P'_{2a} la probabilité pour que la partie dure $2a$ coups; la durée moyenne cherchée est donc

$$\sum_{a=1}^{\infty} 2a.P'_{2a} = 1 + \sum_{a=2}^{\infty} \frac{1\ 3\ 5 \ldots (2a-3)}{2\ 4\ 6 \ldots (2a-2)}$$

$$= 1 + \frac{1}{2} + \frac{1}{2}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6} + \ldots$$

Calculons de proche en proche les sommes des 2, 3, 4, ... premiers termes de cette série; nous avons

$$S_4 = 1 + \frac{1}{2} = \frac{3}{2},$$

$$S_6 = \frac{3}{2} + \frac{1}{2}\ \frac{3}{4} = \frac{3}{2}\cdot\frac{5}{4},$$

$$S_8 = \frac{3}{2}\cdot\frac{5}{4} + \frac{1}{2}\cdot\frac{3}{4}\ \frac{5}{6} = \frac{3}{2}\cdot\frac{5}{4}\ \frac{7}{6}, \qquad \ldots,$$

et, d'une manière générale,

$$S_{2a} = \frac{3}{2}\cdot\frac{5}{4}\ \frac{7}{6}\cdots\frac{2a-1}{2a-2}.$$

S_{2a} augmentant indéfiniment avec a, la durée moyenne de la partie est infinie. On voit ainsi qu'un joueur, qui décide d'arrêter le jeu dès que la partie lui procurera un gain, n'est pas du tout sûr de gagner.

Ceci fournit la réponse à un paradoxe connu, qui résultait de la croyance, que l'on avait, que le joueur, avec cette méthode, a une certitude de gagner. Dans ces conditions, son partenaire pouvait avoir la même certitude, en prenant la même décision, ce qui était évidemment paradoxal.

Si l'on décide d'arrêter le jeu soit à la première égalité, soit, tout au moins, au $(2a)^e$ coup, la durée moyenne des parties sera S_{2a}; sa valeur est asymptotiquement égale à $2\sqrt{\frac{a}{\pi}}$, si a est très grand. Par exemple, si la partie comporte 10000 coups au plus, sa durée moyenne est de l'ordre de 80 coups.

6. **Problème.** — *Étant donnée une partie de 2p coups de pile ou face, quel est le nombre probable de retours à l'égalité?*

Les retours à l'égalité sont les points de rencontre, avec OA, du chemin représentatif de la partie. Soit B un point de rencontre, de coordonnées (a, a); la probabilité pour que le contour passe par B est

$$P_{2a} = \frac{1.3.5\ldots 2a-1}{2.4.6\ldots 2a}.$$

Si l'on fait n parties, le nombre probable de contours qui passeront par B est nP_{2a}. Si chaque partie comporte $2p$ coups, a peut prendre toutes les valeurs de 0 à p, et le nombre total de points d'intersection de OA par les chemins représentatifs de ces n parties est

$$n\sum_{a=0}^{p} P_{2a}$$

Le point O est ainsi compté n fois, le nombre moyen cherché est enfin

$$\sum_{a=0}^{p} P_{2a} = 1 + \frac{1}{2} + \frac{1}{2}\cdot\frac{3}{4} + \ldots + \frac{1.3.5\ldots 2p-1}{2.4.6\ldots 2p} = \frac{3.5\ldots(2p+1)}{2.4\ldots 2p}.$$

Ce nombre est égal à la longueur moyenne S_{2p+2} des parties, de $2p+2$ coups au plus, quand on les arrête à la première égalité. Il est équivalent à $2\sqrt{\frac{p}{\pi}}$, lorsque p est infiniment grand.

La densité des points d'intersection, sur la droite OA, est alors de l'ordre de $\frac{2}{p}\sqrt{\frac{p}{\pi}} = \frac{2}{\sqrt{\pi p}}$; elle tend vers zéro avec $\frac{1}{p}$. Ce résultat est une confirmation du fait que la durée moyenne de la partie est infinie, lorsque l'on convient de ne s'arrêter qu'à la première égalité.

7. Jeu de pile ou face avec mise limitée. — Dans tout ce qui précède, les joueurs étaient susceptibles de perdre une somme quelconque. En réalité, la fortune d'un joueur est limitée, et il entreprend une partie avec une mise encore plus limitée, en général. Il est donc obligé de s'arrêter quand sa mise est perdue.

Le schéma du quadrillage va nous permettre encore d'étudier une partie, ainsi entreprise. On peut choisir l'origine de la partie, de façon que son extrémité soit encore sur OA; il suffit en effet de

prendre pour origine, non le point O, mais un point B situé sur l'axe des y négatifs.

Si la mise du joueur est égale à n fois l'enjeu de chaque coup, le point origine B aura les coordonnées $(0, -n)$; dans ces conditions, la partie s'arrêtera en O si le joueur perd les n premiers coups, et s'arrêtera, d'une manière générale, en un point de OA, dès que le nombre de coups perdus surpassera de n unités le nombre de coups gagnés

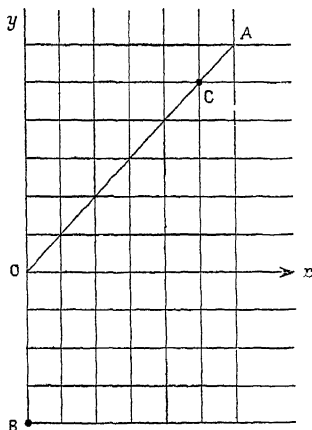



Fig. 18.

Ceci posé, cherchons la probabilité pour que cette mise soit complètement perdue au bout de $(n+2p)$ coups.

Remarquons tout d'abord que la mise n ne peut être perdue qu'au bout d'un nombre de coups de la forme $n+2p$, et que la partie se termine alors au point $C(p, p)$ de la bissectrice OA. La question est la suivante :

Parmi toutes les parties commencées en B, *combien se terminent au point* C *sans avoir rencontré le segment de bissectrice* OC ?

Ou encore, sous une forme qui fait apparaître de suite la solution :

Parmi tous les chemins issus de C, *combien se terminent en* B *sans avoir rencontré* CO ?

On reconnaît là l'énoncé du problème du scrutin, dans lequel on prendrait C comme origine des axes. Les coordonnées de B par

rapport à ces axes sont $(n+p, p)$, et le nombre de ces chemins est alors

$$\frac{(n+2p)!}{p!\,(n+p)!}\,\frac{n}{n+2p}=\frac{n(n+p+1)\ldots(n+2p-1)}{p!}.$$

On sait que le nombre des parties de $n+2p$ coups est 2^{n+2p}, et l'on obtient alors, pour la probabilité en question, la valeur

$$P^n_{2p}=\frac{n(n+p+1)(n+p+2)\ldots(n+2p-1)}{2^{n+2p}\,p!}.$$

Remarque. — Si la mise est égale à l'enjeu, n est égal à 1, et l'on a

$$P^1_{2p}=\frac{(p+2)(p+3)\ldots 2p}{2^{2p+1}\,p!}=\frac{(2p)!}{(2p+2)\,2^{2p}(p!)^2},$$

$$P^1_{2p}=\frac{1.1.3\ldots(2p-1)}{2.4.6\ldots(2p+2)}.$$

C'est la probabilité, que nous avons désignée plus haut par P'_{2p+2}, pour que, partant de O, la première égalité se produise au bout de $2p+2$ coups.

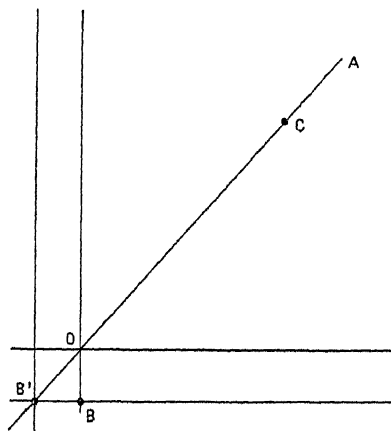



Fig. 19.

Le schéma géométrique rend cette identité évidente, car si B′ est le point de coordonnées $(-1, -1)$, toute partie, d'origine B′, et d'extrémité C, qui reste au-dessous de OA entre ces deux points, passe par B.

8. Un raisonnement analogue permet de déterminer la probabilité

P^n_{2p}, en établissant, pour cette probabilité, une formule récurrente par rapport à n. La remarque que nous venons de faire nous a donné la probabilité P^1_{2p} à partir de l'expression connue de la probabilité P'_{2a}. Nous pouvons donc supposer que la formule

$$P^n_{2p} = \frac{n(n+p+1)\ldots(n+2p-1)}{2^{n+2p}p!},$$

établie pour $n=1$, p quelconque, est vérifiée jusqu'à une certaine valeur de n; il suffit de montrer qu'elle est encore valable pour la valeur suivante de n.

Soient B le point représentatif de la mise n, B′ celui de la mise $n+1$. Un chemin, issu de B′, a autant de chance de passer par B que de passer par le point B_1, de coordonnées $(1, -n)$; s'il passe par B,

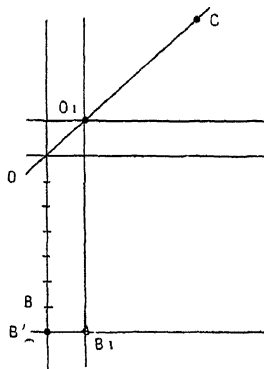



Fig. 20.

c'est un chemin de mise n; s'il passe par B_1, il devient chemin de mise $n+2$; nous avons donc la relation

$$2P^{n+1}_{2p} = P^n_{2p} + P^{n+2}_{2p-2}$$

ou

$$P^{n+2}_{2p-2} = 2P^{n+1}_{2p} - P^n_{2p}.$$

Si l'on admet que la formule, dont nous voulons démontrer la généralité, est exacte pour n et $n+1$, on a bien

$$P^{n+2}_{2p-2} = \frac{(n+1)(n+p+2)\ldots(n+2p)}{2^{n+2p}p!} - \frac{n(n+p+1)\ldots(n+2p-1)}{2^{n+2p}p!}$$

$$= \frac{(n+2)(n+p+2)\ldots(n+2p-1)}{2^{n+2p}(p-1)!}.$$

9. Valeur asymptotique de P^n_{2p}. — Proposons-nous maintenant de déterminer la valeur asymptotique de la probabilité

$$P^n_{2p} = \frac{n}{n+2p}\,\frac{(n+2p)!}{p!\,(n+p)!\,2^{n+2p}},$$

lorsque la mise n, et le nombre $2p$ de coups sont très grands, ce dernier nombre étant, de plus, infiniment grand par rapport au premier.

Il suffit d'appliquer la formule asymptotique de Stirling

$$\log m! = \left(m+\frac{1}{2}\right)\log m - m + \frac{1}{2}\log 2\pi,$$

ce qui donne

$$\begin{aligned}
\log P^n_{2p} = &\left(n+2p+\frac{1}{2}\right)\log(n+2p) - (n+2p) + \frac{1}{2}\log 2\pi,\\
&-\left(p+\frac{1}{2}\right)\log p \qquad\quad + p \qquad\quad - \frac{1}{2}\log 2\pi,\\
&-\left(n+p+\frac{1}{2}\right)\log(n+p) + (n+p) - \frac{1}{2}\log 2\pi,\\
&+\log n - \log(n+2p) \qquad - (n+2p)\log 2
\end{aligned}$$

Or, d'après l'hypothese faite sur l'ordre de grandeur de n par rapport à p, nous pouvons écrire

$$\begin{aligned}
\log(n+p) &= \log p + \log\left(1+\frac{n}{p}\right) = \log p \qquad\quad + \frac{n}{p} - \frac{n^2}{2p^2} + \ldots,\\
\log(n+2p) &= \log 2p + \log\left(1+\frac{n}{2p}\right) = \log 2 + \log p + \frac{n}{2p} - \frac{n^2}{8p^2} + \ldots;
\end{aligned}$$

et il vient alors

$$\begin{aligned}
\log P^n_{2p} = &-\frac{3}{2}\log p - \frac{1}{2}\log 2 - \frac{1}{2}\log 2\pi + \log n\\
&+\left(n+2p-\frac{1}{2}\right)\left(\frac{n}{2p}-\frac{n^2}{8p^2}+\ldots\right) - \left(n+p+\frac{1}{2}\right)\left(\frac{n}{p}-\frac{n^2}{2p^2}+\ldots\right).
\end{aligned}$$

Si nous remarquons que $\frac{n}{p}$ est négligeable devant le terme en $\frac{n^2}{p}$, la partie principale du second membre donne

$$\log P^n_{2p} = \log\frac{n}{2p^{\frac{3}{2}}\sqrt{\pi}} - \frac{n^2}{4p},$$

et par suite, pour valeur asymptotique de la probabilité elle-même,

$$P^n_{2p} = \frac{n}{2p\sqrt{p\pi}} e^{-\frac{n^2}{4p}}$$

Si p est infiniment grand, non seulement par rapport à n, mais encore par rapport à n^2, l'exponentielle est équivalente à l'unité, et P^n_{2p} est équivalent à $\frac{n}{2p\sqrt{p\pi}}$.

On voit encore que, quelque grande que soit la mise initiale n, cette probabilité tend vers o quand p augmente indéfiniment. Autrement dit, la probabilité pour que la partie ne s'arrête jamais est nulle

10. On peut se demander quelle est la probabilité, pour la partie, de durer $n+2p$ coups au moins, la mise étant toujours n, p et n sont supposés tels que P^n_{2p} soit équivalent à $\frac{n}{2p\sqrt{p\pi}}$; la probabilité en question est évidemment mesurée par la somme de la série

$$\sum_{q=p}^{\infty} \frac{n}{2q\sqrt{\pi q}}.$$

La courbe $y = \frac{n}{2x\sqrt{\pi x}}$ étant asymptotique à Ox, et p étant très grand, on peut évaluer approximativement la somme de cette série en la remplaçant par l'intégrale définie

$$\int_p^{\infty} \frac{n\,dq}{2q\sqrt{\pi q}} = \frac{n}{\sqrt{\pi p}}.$$

En particulier, pour la valeur 1 de n, nous retrouvons la valeur asymptotique de la probabilité Q'_{2p} pour qu'une partie, commencée au point O, ne se termine pas avant $2p+2$ coups.

Remarquons enfin que lorsque p est petit, l'influence de n dans P^n_{2p} est de l'ordre de $\frac{1}{2^n}$, et que, par contre, si p est très grand, ainsi que n, cette influence est de l'ordre de n.

Appliquons ces résultats à l'exemple numérique suivant : Supposons que n soit égal à 100, et $2p$ à 500 000.

La probabilité correspondante a pour valeur $\frac{1}{5\sqrt{\pi}}$, ou, approximativement, $\frac{1}{10}$.

Le joueur n'a qu'une chance sur dix, environ, de ne perdre sa mise qu'après 500 000 coups. Sa ruine serait d'autant plus lente que sa mise n serait plus forte, autrement dit qu'il jouerait un jeu plus petit

11. Dans ces conditions, on peut se demander si un joueur, qui dispose de deux mises, a avantage à les engager dans deux parties distinctes, ou doit, au contraire, les associer dans une seule partie

La probabilité pour qu'il soit ruiné avant $n+2p$ coups, avec la mise n, est $1-\alpha$, en désignant par α la quantité $\frac{n}{\sqrt{\pi p}}$. Si n' est la seconde mise dont il dispose, il y a une probabilité $1-\alpha' = 1-\frac{n'}{\sqrt{\pi p}}$ pour qu'il la perde avant $n'+2p$ coups.

La probabilité pour qu'il perde ses deux mises dans ces deux parties est donc $(1-\alpha)(1-\alpha')$ ou, sensiblement, $1-\alpha-\alpha'$, car le produit $\alpha\alpha'$ est négligeable.

S'il jouait, dans une même partie, la mise $n+n'$, la probabilité de la perdre avant le même nombre total de coups, c'est-à-dire $n+n'+4p$ coups, serait $1-\frac{n+n'}{\sqrt{2\pi p}}$, ou encore $1-\frac{\alpha+\alpha'}{\sqrt{2}}$, plus grande que la précédente.

Avec l'exemple numérique précédent, prenons une somme égale à $1000q$, q étant un nombre entier. Si l'on fait $10q$ parties de 500 000 coups au plus, avec, pour chaque partie, la mise initiale 100, la probabilité totale d'être ruiné sera

$$\left(1-\frac{1}{10}\right)^{10q}, \text{ ou, sensiblement, } e^{-q}.$$

Elle est donc très faible. Pour qu'elle soit inférieure à $\frac{1}{1000}$, il suffit que q soit supérieur à 3 Log 10, c'est-à-dire 6 ou 7. On voit ainsi qu'avec une somme égale à 6000 ou 7000 fois l'enjeu, et des parties limitées à 500 000 coups chacune, un joueur a 999 chances sur 1000 de ne pas se ruiner complètement.

Par contre, avec une mise égale à la somme globale, ce joueur a 1 chance sur 7 de se ruiner avant la fin d'une partie de 30 millions de coups. Il a donc tout intérêt à fractionner sa mise, s'il veut retarder sa ruine.

CHAPITRE VI.

STATISTIQUE.

I — GÉNÉRALITÉS SUR LES FONCTIONS DE LA STATISTIQUE

1. Jusqu'ici nous ne nous sommes pas occupés du rapport qui peut exister entre la théorie des probabilités, et les résultats fournis par la réalité. C'est ainsi que la loi de Gauss a été établie indépendamment de toute statistique, et, d'ailleurs, elle ne concorde pas exactement avec les lois des phénomènes naturels

Nous nous proposons, dans ce Chapitre, de faire l'étude des probabilités considérées comme fournies par l'expérience, et non plus, comme précédemment, considérées comme fournies *a priori*

Tout d'abord, il est intéressant de remarquer que ces probabilités sont établies par des statistiques, nécessairement relatives au passé Il semble, à première vue, qu'il y ait là une anomalie, puisque la probabilité ne peut avoir de signification que s'il s'agit d'événements non encore produits, et seulement susceptibles de se produire; mais il est clair que l'on ne peut songer à tirer des conclusions relatives a l'avenir que grâce à des observations du passé, et cette anomalie ne doit pas paraître plus extraordinaire que la recherche et l'application des lois dans les sciences physiques.

Pour bien comprendre en quoi consiste la statistique, et la nature des problèmes qu'elle conduit à se poser, examinons de suite un exemple particulier. Nous choisirons la table de mortalité.

Étant donnés 1000 individus, déterminons les nombres de ceux qui meurent aux différents âges, en les groupant par périodes quinquennales.

On constate que 117 de ces individus meurent avant l'âge de 5 ans, 34 entre 5 et 10 ans, etc., et, d'une manière générale, on peut former

le tableau suivant :

Age ..	5	10	15	20	25	30	35	40	45	50	55
Nombre...	117	34	17	25	26	25	26	28	32	37	45
Age ..	60	65	70	75	80	85	90	95	100		
Nombre ..	57	73	91	97	110	92	54	11	2		

La courbe de variation de ces nombres en fonction de l'âge se présente donc sous une forme compliquée. La vie moyenne, qui est l'âge dont l'abscisse partage en deux parties égales l'aire limitée par cette courbe, ne correspond à aucun point particulier.

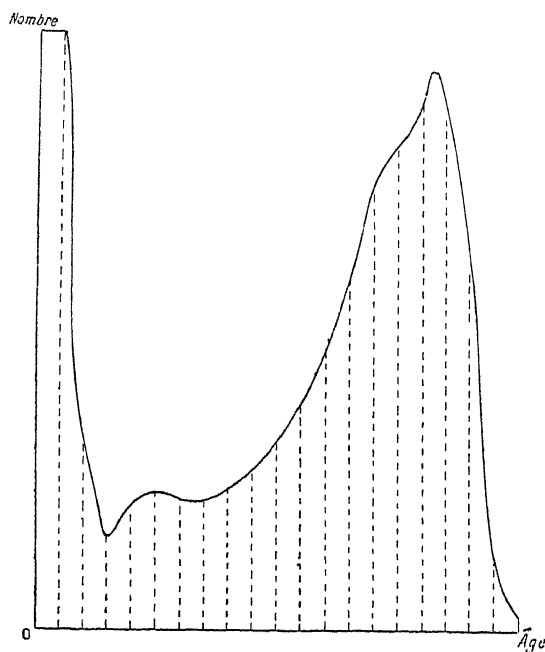



Fig. 21.

On peut donner à cette courbe représentative la forme en escalier; l'aire totale limitée par cette ligne polygonale est alors égale à 1000, nombre des individus, à condition de prendre ici comme unité des abscisses la période de 5 ans. Les probabilités de mortalité sont :

Avant 5 ans..................	0,117
Entre 5 et 10 ans............	0,034
Entre 10 et 15 ans...........	0,017

de sorte que la ligne polygonale des probabilités se déduirait de la courbe précédente par une simple réduction des ordonnées dans la proportion de 1000 à 1, et l'aire limitée par cette ligne aurait pour mesure l'unité.

Malgré sa bizarrerie, cette courbe permet d'en déduire une autre, d'allure beaucoup plus générale, il suffit de considérer les probabilités de mort avant les différents âges.

Par exemple, la probabilité de mourir avant 10 ans est

$$0,117 + 0,034 = 0,151,$$

avant 15 ans, c'est $0,151 + 0,017 = 0,168$, etc., et, d'une manière générale, on obtient les résultats rassemblés dans le tableau suivant

Age	Probabilité	Age	Probabilité.
5	0,117	55	0,412
10	0,151	60	0,469
15	0,168	65	0,542
20	0,193	70	0,633
25	0,219	75	0,730
30	0,244	80	0,840
35	0,270	85	0,932
40	0,298	90	0,986
45	0,330	95	0,988
50.	0,367	100	1,000

La courbe représentative de ce tableau est alors croissante, de 0 à 1. On conçoit immédiatement que l'allure sera analogue pour toutes les courbes de probabilité construites de cette façon.

2. D'une manière générale, on est conduit à considérer des fonctions $f(x)$, définies pour toutes les valeurs de la variable x, supposée varier de $-\infty$ à $+\infty$, une telle fonction $f(x)$ représentant la probabilité pour qu'une certaine quantité variable z prenne une valeur inférieure ou égale à x.

Il est clair qu'une pareille fonction doit vérifier *a priori* certaines conditions; il faut, tout d'abord, qu'elle soit nulle pour $x = -\infty$, et égale à l'unité pour $x = +\infty$; d'autre part, c'est une fonction nécessairement non décroissante. Mais $f(x)$ peut être discontinue; c'est ce qui se présente, par exemple, si la quantité variable z est elle-même discontinue.

Prenons comme exemple simple de discontinuité le jeu de pile ou face, en désignant par z le gain résultant d'un coup de dé ; si 1 est l'enjeu, on a la probabilité $\frac{1}{2}$ de gagner $+1$, et la même probabilité de gagner -1. On a donc

$$
\begin{aligned}
f(x) &= 0 \quad \text{pour} \quad x < -1, \\
f(-1) &= \frac{1}{2} \\
f(x) &= \frac{1}{2} \quad \text{pour} \quad -1 < x < 1, \\
f(1) &= 1 \\
f(x) &= 1 \quad \text{pour} \quad x > 1
\end{aligned}
$$

La courbe représentative $z = f(x)$ est la courbe en escalier $-\infty$ ABCD ∞.

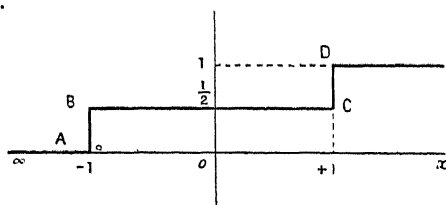



Fig. 22.

On obtient une courbe analogue pour représenter le gain de deux coups consécutifs ; dans cet exemple, la variation de z est limitée à l'intervalle $(-2, +2)$, et peut se résumer dans le tableau suivant :

x	$-\infty$	-2		-1		0		$+1$		$+2$		$+\infty$
$f(x)$	0	$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$		$\frac{3}{4}$	$\frac{3}{4}$	1	

Sa courbe représentative contient une marche de plus que la précédente :

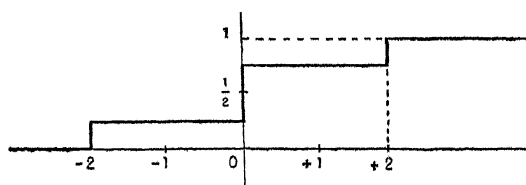



Fig. 23.

De telles fonctions sont, soit constantes, soit brusquement croissantes d'une quantité finie. Il est clair qu'en un point de disconti-

nuité, on doit considérer cette discontinuité comme existant à gauche du point et non à droite. C'est ainsi que, sur la deuxième courbe, on avait

$$f(x) = \frac{3}{4} \quad \text{pour} \quad 2 > x \geqq 0$$

et

$$f(x) = \frac{1}{4} \quad \text{pour} \quad -2 \leqq x < 0.$$

3. Pour entreprendre l'étude des fonctions de cette nature, nous ferons appel à la notation de Stieltjes.

Désignons par $df(x)$ l'accroissement de $f(x)$ dans un intervalle infiniment petit dx; c'est encore la probabilité pour que la quantité variable z soit comprise dans l'intervalle dx. Si $f(x)$ admet une dérivée, $df(x)$ désigne la différentielle $f'(x)dx$ de cette fonction, et l'intégrale $\int_a^b df(x) = \int_a^b f'(x)\,dx = f(b) - f(a)$ mesure l'accroissement de $f(x)$ entre a et b.

D'une manière générale, $\int_a^b df(x)$ représentera cet accroissement $f(b) - f(a)$, même lorsque l'intégrale intermédiaire $\int_a^b f'(x)\,dx$ n'aura aucun sens.

La différentielle $df(x)$ est la probabilité élémentaire, car la probabilité pour que la quantité variable z soit comprise entre a et b $(a < z \leqq b)$ est

$$A = f(b) - f(a) = \int_a^b df(x).$$

Si toutes les valeurs de z sont comprises entre a et b, on a $A = 1$.

Dans le cas des courbes discontinues, il peut exister une probabilité pour que z prenne une valeur déterminée; cette probabilité est, en effet, mesurée par l'accroissement brusque de $f(x)$ pour cette valeur particulière. Si la courbe est continue, on ne peut parler que de probabilité pour que z soit compris entre deux valeurs. Dans le cas où la courbe représentative est continue, non décroissante entre les valeurs 0 et 1, et asymptote à l'axe des x négatifs, et à la droite $z = 1$, la probabilité est définie entre $-\infty$ et $+\infty$ et $\int_{-\infty}^{+\infty} df(x) = 1$ se présente sous la forme d'une intégrale ordinaire (*fig.* 24).

Mais on peut avoir la combinaison d'une courbe continue dans certains intervalles avec des ordonnées verticales en certains points. On peut aussi avoir des ensembles parfaits

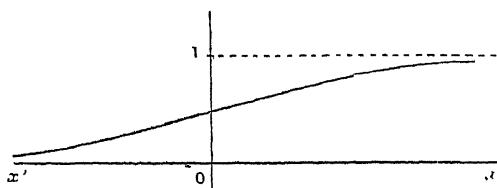



Fig. 24

4. La notation de Stieltjes permet encore de considérer des intégrales de la forme $\int_a^b g(x)\,df(x)$, dans lesquelles $g(x)$ est une fonction quelconque de x, définie ainsi que $f(x)$, dans l'intervalle $(-\infty, +\infty)$. Stieltjes appelle ces intégrales des « moments »; il considère, en effet, la fonction $f(x)$ comme définissant une distribution de masse sur l'axe des x; $df(x)$ représente la masse placée au point x, si $f(x)$ est discontinue en ce point; si, au contraire, $f(x)$ est continue et admet une dérivée en un point, cette dérivée représente la densité de la matière en ce point.

On a ainsi une répartition de matière le long de l'axe des x, dont la masse totale est égale à l'unité, d'après l'hypothèse faite sur la valeur de l'intégrale $\int_{-\infty}^{+\infty} df(x)$, et l'on voit que cette répartition peut être continue, sauf en certains points isolés où peuvent être accumulées des masses finies.

Ceci posé, il est naturel d'appeler moments des divers ordres, par rapport à l'origine o, les sommes $\sum_i m_i x_i$, $\sum_i m_i x_i^2$, $\sum_i m_i x_i^3$, ..., étendues à toute la masse répartie sur l'axe $x'x$; or, ces sommes, d'après la définition même de l'intégrale de Stieltjes, peuvent être représentées par des intégrales de la forme $\int_a^b x\,df(x)$, $\int_a^b x^2\,df(x)$, $\int_a^b x^3\,df(x)$, ..., et cela, sans être obligé de faire aucune hypothèse sur la continuité de la répartition de ces masses.

On peut encore définir, d'une manière précise, l'intégrale de

Stieltjes $\int_a^b x^n df(x)$, en employant le même raisonnement que pour les intégrales définies ordinaires; on peut supposer l'intervalle (a, b) partagé par des points de division $x_1, x_2, \ldots, x_p$; en désignant par ξ_i un point arbitraire de l'intervalle partiel $x_i x_{i+1}$, on forme la somme $\sum_i \xi_i^n [f(x_{i+1}) - f(x_i)]$, et l'on fait tendre vers zéro tous les intervalles partiels; la valeur de l'intégrale de Stieltjes peut être définie comme la limite de cette somme, si cette limite existe.

Mais il ne faut pas oublier qu'ici, $f(x_{i+1}) - f(x_i)$ ne tend pas nécessairement vers zéro avec $x_{i+1} - x_i$. Si, en un point d'abscisse c, on a une masse finie m_c, et si x_i et x_{i+1} tendent vers c, tout en restant de part et d'autre, la limite, en ce point, de $\xi_i^n [f(x_{i+1}) - f(x_i)]$ est $c^n m_c$.

Nous supposerons, dans ce qui suit, que l'intégrale $\int_{-\infty}^{+\infty} x^n df(x)$ a un sens quel que soit l'entier positif n, ceci exige, en particulier, que la valeur de $x^n df(x)$, relative à un intervalle de longueur donnée, tende vers zéro lorsque x augmente indéfiniment; par exemple, il faut avoir $\lim_{x=\infty} x^n [f(x+1) - f(x)] = 0$.

Nous désignerons par μ_n le moment $\int_{-\infty}^{+\infty} x^n df(x)$, en particulier, $\mu_0 = \int_{-\infty}^{+\infty} df(x)$ représente la masse totale repartie, et, par hypothèse, est égal à l'unite $\mu_1 = \int_{-\infty}^{+\infty} x\, df(x)$ représente l'abscisse du point, par rapport auquel le moment de la masse est nulle, en effet, on a $\int_{-\infty}^{+\infty} (x - \mu_1) df(x) = 0$, la droite, parallèle à l'axe des y, menée par ce point, s'appelle la « droite médiane »; μ_1 est nul si on la choisit pour axe des y.

Un cas intéressant est celui des courbes qui admettent, sur l'axe des y, un centre de symétrie; pour ces courbes, tous les moments d'ordre impair sont nuls.

Les moments d'ordre pair, qui ne peuvent jamais être nuls, représentent encore les valeurs moyennes des puissances deuxième, quatrième, sixième, etc., de z.

5. Les masses $df(x)$ peuvent être réparties d'une manière continue

sans toutefois admettre partout une densité de répartition. Cependant, on peut observer qu'il résulte, de la définition même de la fonction non décroissante $f(x)$, que, en un point de discontinuité, la discontinuité existera toujours avant, sans exister nécessairement après, il se produit quelque chose d'analogue à ce que nous avons déjà observé sur les courbes en escalier; par exemple, la courbe peut avoir une dérivée à droite du point de discontinuité, alors que la dérivée à gauche serait infinie.

On peut même être conduit à considérer des fonctions continues, non décroissantes, analogues aux fonctions en escalier, et dont la dérivée n'existe pas en tous les points de l'intervalle où la fonction est définie. Examinons, tout d'abord, quelques fonctions de cette nature, définies dans un intervalle fini $(0, 1)$ par exemple, et constantes dans une infinité d'intervalles partiels intérieurs à cet intervalle. Ces intervalles partiels peuvent avoir une étendue totale égale à la mesure de l'intervalle $(0, 1)$ ou inférieure à l'unité.

6. **Premier exemple.** — Voici, tout d'abord, un exemple où la mesure des intervalles, dans lesquels $f(x)$ a une valeur constante, est égale à 1. On peut supposer que $f(x)$ croît de zéro à 1 dans l'intervalle $(0, 1)$, et est constante en dehors de cet intervalle, de sorte que l'on a

$$f(x) = 0 \qquad \text{pour} \qquad x < 0,$$
$$f(x) = 1 \qquad \text{pour} \qquad x \geq 1.$$

Pour définir $f(x)$, partageons l'intervalle $(0, 1)$ en trois parties égales, et donnons à cette fonction la valeur $\frac{1}{2}$ dans l'intervalle

$$\frac{1}{3} \leq x < \frac{2}{3};$$

partageons ensuite chacun des deux autres intervalles $\left(0, \frac{1}{3}\right)$ et $\left(\frac{2}{3}, 1\right)$ en trois parties égales, et supposons $f(x)$ constante dans les portions médianes :

$$f(x) = \frac{1}{4} \qquad \text{pour} \qquad \frac{1}{9} \leq x < \frac{2}{9},$$
$$f(x) = \frac{3}{4} \qquad \text{pour} \qquad \frac{7}{9} \leq x < \frac{8}{9}.$$

Partageons encore chacun des quatre intervalles restants en trois

parties égales, et, dans les quatre intervalles médians, donnons à $f(x)$ les valeurs constantes suivantes :

$$f(x)=\frac{1}{8} \quad \text{pour} \quad \frac{1}{27} \leq x < \frac{2}{27},$$

$$f(x)=\frac{3}{8} \quad \text{pour} \quad \frac{7}{27} \leq x < \frac{8}{27},$$

$$f(x)=\frac{5}{8} \quad \text{pour} \quad \frac{19}{27} \leq x < \frac{20}{27},$$

$$f(x)=\frac{7}{8} \quad \text{pour} \quad \frac{25}{27} \leq x < \frac{26}{27},$$

et ainsi de suite

Dans ces conditions, la fonction $f(x)$ est constante dans tout intervalle où elle est définie.

Dans la première opération, elle est définie dans un intervalle de mesure $\frac{1}{3}$; dans la deuxième, on la définit dans deux intervalles de mesure totale $\frac{2}{3^2}$, ensuite, on la definit dans quatre intervalles de mesure totale $\frac{2^2}{3^3}$, etc., la mesure de tous les intervalles partiels, compris entre 0 et 1, dans lesquels $f(x)$ sera définie, est donc

$$\frac{1}{3}+\frac{2}{3^2}+\frac{2^2}{3^3}+\ldots+\frac{2^n}{3^{n+1}}+\ldots=1$$

Il nous reste à donner une définition systématique de cette fonction dans l'intervalle $(0, 1)$

Remarquons, pour cela, que les points de division de cet intervalle, le partageant successivement en 3, 3^2, 3^3, .. parties égales, peuvent être représentés simplement dans le système de numération ternaire. Supposons que x soit exprimé dans ce système de numération

$$x=\frac{\alpha_1}{3}+\frac{\alpha_2}{3^2}+\frac{\alpha_3}{3^3}+\ldots+\frac{\alpha_n}{3^n}+\ldots$$

où $\alpha_1, \alpha_2, \ldots, \alpha_n$ sont les chiffres 0, 1, 2. On écrira encore

$$x=0_3, \alpha_1\alpha_2\alpha_3\ldots\alpha_n\ldots;$$

l'indice 3 qui accompagne la partie entière indique la base de la numération.

Reprenons, avec cette notation, la définition de notre fonction. Tout d'abord, nous avons défini $f(x)$ dans l'intervalle $0_3, 1 \leq x < 0_3, 2$, sa

valeur dans cet intervalle étant, exprimée en numération binaire, $f(x) = 0_2, 1$.

Après le deuxième partage, on définit $f(x)$ simultanément dans l'intervalle $0_3, 01 \leqq x < 0_3, 02$, dans lequel $f(x) = 0_2, 01$, et dans l'intervalle $0_3, 21 \leqq x < 0_3, 22$, dans lequel $f(x) = 0_2, 11$, et ainsi de suite

Sans aller plus loin, on voit que les intervalles partiels dans lesquels $f(x)$ sera successivement définie sont l'ensemble des valeurs de x, qui, en numération ternaire, contiennent le chiffre 1 au moins une fois. Si le premier chiffre 1 de ce nombre est le $k^{\text{ième}}$ chiffre, à partir de la virgule, la valeur correspondante de $f(x)$ sera définie dans la $k^{\text{ième}}$ opération.

Il résulte de là que les seules valeurs de x pour lesquelles $f(x)$ ne sera pas définie, seront celles qui peuvent ne s'écrire qu'avec les seuls chiffres 0 et 2. Remarquons que la probabilité pour qu'un nombre ne contienne que ces deux seuls chiffres, en numération ternaire, est $\lim\limits_{n=\infty} \left(\frac{2}{3}\right)^n = 0$, il n'en est pas moins vrai qu'il y a une infinité de tels points, mais, d'après ce qui précède, nous savons qu'ils forment un ensemble de mesure nulle

Cependant $\frac{2}{3}$ s'écrit aussi bien $0_3, 2$ ou $0_3, 1111\ldots$, avec une infinité de chiffres 1 successifs, sans que $f(x)$ soit définie pour $x = \frac{2}{3}$; il se présente ici une légère difficulté, due à ce que la numération ternaire, comme d'ailleurs toutes les numérations, n'est pas une représentation univoque; elle sera d'ailleurs aisée à surmonter.

La comparaison des valeurs de $f(x)$ aux extrémités gauches des intervalles partiels dans lesquels cette fonction est définie, permet de former le tableau suivant :

$$
\begin{aligned}
f(0_3,\ 1) &= 0_2,\ 1,\\
f(0_3,\ 0\ 1) &= 0_2,\ 0\ 1,\\
f(0_3,\ 2\ 1) &= 0_2,\ 1\ 1,\\
f(0_3,\ 0\ 0\ 1) &= 0_2,\ 0\ 0\ 1,\\
f(0_3,\ 0\ 2\ 1) &= 0_2,\ 0\ 1\ 1,\\
f(0_3,\ 2\ 0\ 1) &= 0_3,\ 1\ 0\ 1,\\
f(0_3,\ 2\ 2\ 1) &= 0_2,\ 1\ 1\ 1,\\
\ldots\ \ldots &\ldots\ldots\ldots\ldots
\end{aligned}
$$

On constate sans difficulté qu'on passe de la valeur de x à celle

de $f(x)$ en substituant la numération binaire à la numération ternaire, et en remplaçant tous les chiffres 2 par le chiffre 1

En se rappelant que, pour un nombre x qui contient des chiffres après le premier chiffre 1, $f(x)$ a la même valeur que pour le nombre déduit de l'expression de x par suppression de tous les chiffres à droite de ce chiffre 1, on peut énoncer enfin la règle suivante

« La fonction $f(x)$ est définie pour tous les nombres x qui contiennent le chiffre 1, en numération ternaire (exception faite de ceux qui contiennent une succession infinie de 1). Pour définir la valeur de $f(x)$ en un tel point, on se borne au premier chiffre 1 de x, et l'on remplace les chiffres 2 du nombre obtenu par des 1 Ce dernier nombre représente alors $f(x)$ dans le système binaire »

Cette règle, quelque bizarre qu'elle paraisse, fournit une définition arithmetique précise de $f(x)$, dans certains intervalles Nous pourrons en déduire la définition de cette fonction, pour toutes les valeurs de x, si nous l'assujettissons à être non decroissante, ou, ce qui revient au même, continue. Montrons de suite que ces deux conditions sont equivalentes

En effet, après le $n^{\text{ième}}$ partage, l'oscillation de la fonction non décroissante $f(x)$, dans un intervalle de mesure $\frac{1}{3^n}$, dans lequel cette fonction n'est pas definie, est égale à $\frac{1}{2^n}$.

Un nombre x, qui ne contient que des 0 et des 2, peut contenir une infinité de zéros successifs, dans ces conditions, il fait partie de la categorie des nombres fractionnaires, on voit de suite, par continuité, que $f(x)$ s'obtient en remplaçant les chiffres 2 par les 1. Il peut egalement contenir une succession infinie de 2; il suffit de supprimer tous les chiffres 2 successifs, en remplaçant le dernier zéro par 1; $f(x)$ a éte définie pour ces valeurs de x, qui sont les extrémités inférieures des intervalles de definition [1].

Enfin, x peut contenir une infinité de 0 et de 2. Nous allons voir que la valeur de $f(x)$ s'obtient pour ces nombres d'après la règle établie plus haut. En effet, un tel nombre x contient nécessairement

---

(1) Remarquons que, pour ces valeurs de x, $f(x)$ s'obtient par l'application de la règle générale, quelle que soit l'écriture adoptée.

une infinité de fois le chiffre 2 suivi d'un 0. Écrivons, par exemple,

$$x = 0_3,\ 0\ 2\ 2\ 2\ 0\ 2\ \ldots\ 2\ 0\ \ldots\ 2\ 0\ \ldots,$$

et considérons les deux nombres fractionnaires

$$x_1 = 0_3,\ 0\ 2\ 2\ 2\ 0\ 2\ \ldots\ 1$$

et

$$x'_1 = 0_3,\ 0\ 2\ 2\ 2\ 0\ 2\ \ldots\ 2\ 1,$$

le premier est inférieur à x, et le second lui est supérieur. Or on a

$$f(x_1) = 0_2,\ 0\ 1\ 1\ 1\ 0\ 1\ \ldots\ 1$$

et

$$f(x'_1) = 0_2,\ 0\ 1\ 1\ 1\ 0\ 1\ \ldots\ 1\ 1,$$

et $f(x)$ est compris entre ces deux valeurs. Si le groupe 2 0 que l'on a remplacé par 1 0 pour former x_1, et par 2 1 pour former x'_1, occupe les rangs $n-1$, n dans le nombre x, la différence

$$f(x'_1) - f(x_1) = \frac{1}{2^n}.$$

En opérant sur les groupes 2 0 successifs, on est conduit à former deux suites de nombres

$$x_1 < x_2 < x_3 < \ldots < x$$
$$x'_1 > x'_2 > x'_3 > \ldots > x,$$

de limite commune x; les valeurs correspondantes de $f(x)$ sont

$$f(x_1) < f(x_2) < f(x_3) < \ldots$$
$$f(x'_1) > f(x'_2) > f(x'_3) > \ldots$$

qui ont une limite commune bien déterminée, puisque la différence $f(x'_i) - f(x_i)$ tend vers zéro lorsque i augmente indéfiniment. D'après l'hypothèse faite sur $f(x)$, c'est cette limite qui est la valeur de $f(x)$ au point x. $f(x_i)$ s'obtenant en remplaçant par des 1 tous les chiffres 2 du nombre x, limité à un certain nombre de chiffres, la limite $f(x)$ s'obtiendra en effectuant ce remplacement dans le nombre x pris en entier, ce qui donne

$$f(x) = 0_2,\ 0\ 1\ 1\ 1\ 0\ 1\ \ldots\ 1\ 0\ \ldots\ 1\ 0\ \ldots.$$

« Autrement dit, la règle arithmétique qui permet de passer de l'écriture de x à celle de $f(x)$, est absolument générale. »

La fonction $f(x)$, ainsi définie, est continue en tout point, a une dérivée nulle dans tout intervalle de définition, c'est-à-dire en tout point x qui contient un chiffre 1. En un point x, dont la représentation ternaire ne contient que des 0 et des 2, l'oscillation de $f(x)$ dans l'intervalle $(x_i\, x'_i)$, de longueur $\frac{1}{3^n}$, est $\frac{1}{2^n}$, de sorte que le rapport de l'accroissement de $f(x)$ à l'accroissement de x est $\left(\frac{3}{2}\right)^n$, dont la valeur augmente indéfiniment avec n ; en un tel point, la dérivée de $f(x)$ est donc infinie.

En définitive, la fonction y que nous avons ainsi définie, est non décroissante, et continue, de 0 à 1, a chaque valeur de y, écrite en numération binaire, correspond toujours un nombre x qui, en numération ternaire, ne contient que des 0 et des 2. La mesure de ces valeurs particulières de x est nulle, alors que celle des valeurs de y est égale à l'unité

Remarquons enfin que si $df(x)$ représente une masse, la masse répartie en un point dont l'abscisse x ne contient que des 0 et des 2, est infiniment petite, bien que la densité de répartition soit infinie en ce point

7 **Deuxième exemple.** — Nous allons examiner maintenant un exemple dans lequel la fonction $f(x)$ n'est définie que dans des intervalles de mesure totale inférieure à l'unité Et, ici encore, on en déduira la définition de $f(x)$ dans l'intervalle entier (0, 1), grâce a l'hypothèse de la non-décroissance de $f(x)$, ou, ce qui revient au même, de sa continuité.

Enfin, on suppose toujours que $f(x)$ est nulle pour les valeurs négatives de x et est égale à 1 pour $x \geqq 1$.

Ceci posé, portons, de part et d'autre du milieu du segment (0, 1), une longueur $\frac{\alpha_1}{2}$ $(\alpha_1 < 1)$, et donnons à $f(x)$ la valeur constante $\frac{1}{2}$ dans l'intervalle obtenu, A_1B_1, de longueur α_1 ; l'extrémité droite de cet intervalle est exclue de la définition, comme d'ailleurs les extrémités droites de tous les intervalles analogues que nous allons former.

De part et d'autre du milieu de OA_1, portons une longueur égale au $\left(\frac{\alpha_2}{2}\right)^{\text{ième}}$ de OA_1 $(\alpha_2 < 1)$, et effectuons la même opération dans l'intervalle $(B_1, 1)$. Dans l'intervalle A_2B_2, faisons $f(x) = \frac{1}{4}$, et, dans

$A'_2B'_2$, $f(x) = \frac{3}{4}$. Recommençons sur les quatre intervalles OA_2, B_2A_1, $B_1A'_2$, $B'_2 1$, et ainsi de suite, de sorte qu'à la $n^{ième}$ opération, on enleve la $(\alpha_n)^{ième}$ partie de ce qui reste après l'opération precedente

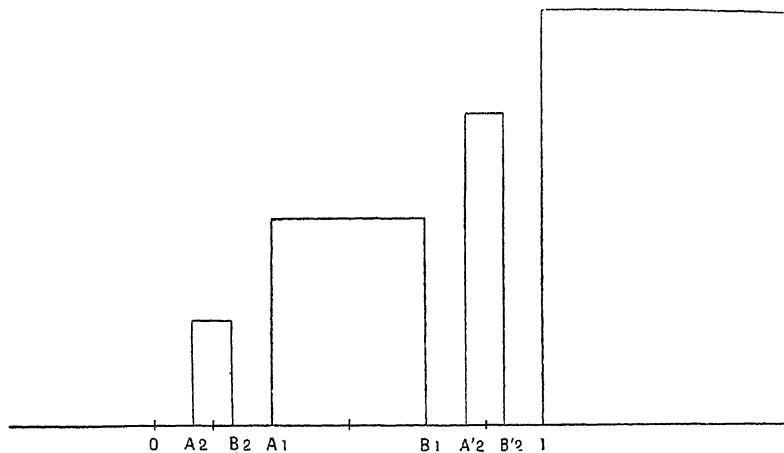



Fig 25

Après la première opération, les deux intervalles restants ont, pour mesure, $1 - \alpha_1$; après la deuxième operation, les quatre intervalles restants mesurent $(1 - \alpha_1)(1 - \alpha_2)$, et, d'une maniere generale, après la $n^{ième}$ opération, il reste 2^n intervalles de mesure totale $\beta_n = (1 - \alpha_1)(1 - \alpha_2)\ldots(1 - \alpha_n)$.

La fonction définie dans le premier exemple n'est qu'un cas particulier de celle-ci, dans lequel tous les α_i ont la valeur commune $\frac{1}{3}$, de sorte que β_n est égal à $\left(\frac{2}{3}\right)^n$. On a alors $\lim\limits_{n=\infty} \beta_n = 0$.

Mais on voit de suite que ceci n'a plus lieu si la serie des α_n est convergente, c'est d'ailleurs la condition nécessaire et suffisante. On peut supposer, par exemple, que $\alpha_n = \frac{1}{2^n}$, ou $\alpha_n = \frac{1}{4n^2}$; dans ce dernier cas, on a

$$\beta = \lim_{n=\infty} \beta_n = \prod_{n=1}^{\infty} \left(1 - \frac{1}{4n^2}\right),$$

ou $\beta = \frac{2}{\pi}$.

Au $n^{ième}$ stade, chacun des 2^n intervalles dans lesquels $f(x)$ n'est

pas définie à la longueur $\frac{\beta_n}{2^n}$; d'autre part, l'oscillation de $f(x)$ entre deux intervalles consécutifs, dans lesquels $f(x)$ est définie, est $\frac{1}{2^n}$; donc la pente de $f(x)$ entre ces deux intervalles est $\frac{1}{\beta_n}$. Il résulte de là que, en un point x de l'ensemble de mesure non nulle dans lequel $f(x)$ ne sera pas directement défini, $f(x)$, étant non décroissante, pourra être définie par une limite. Mais, contrairement à l'exemple précédent, $f(x)$ a ici une dérivée, $\frac{1}{\beta}$, limite de $\frac{1}{\beta_n}$ quand l'intervalle de mesure $\frac{\beta_n}{2^n}$, qui contient x, tend vers zéro.

La fonction ainsi définie a donc une dérivée première en tous les points; mais cette dérivée est discontinue. $df(x)$ représente la répartition continue d'une masse finie, avec une densité discontinue.

Remarque. — Des lois de répartition aussi bizarres ne sont pas seulement curieuses, elles se présentent dans un grand nombre de questions. Des fonctions apparaissent, souvent, comme des limites de fonctions simples, et sont elles-mêmes singulières, et il est intéressant de savoir distinguer ces fonctions des fonctions ordinaires. De telles fonctions se présentent dans un problème très important, le problème des moments, dont l'étude va faire l'objet des paragraphes suivants.

II. — PROBLÈME DES MOMENTS.

8. Considérons un phénomène qui obéit à une certaine loi de probabilité. On mesure des écarts, leur moyenne, ainsi que les moyennes de leurs puissances successives. Peut-on déduire, de ces mesures, la loi de répartition des écarts de ce phénomène?

Supposons, par exemple, que cette loi soit la loi normale des écarts $\varphi_1(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$; on sait que les moyennes des puissances impaires

$$M_{2n+1} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} x^{2n+1} e^{-x^2} \, dx$$

sont nulles; désignons par

$$M_{2n} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} x^{2n} e^{-x^2} \, dx$$

les moyennes des puissances paires.

Inversement, si les moyennes relatives à un phénomène donné tendent vers ces nombres M_k, lorsque le nombre des expériences faites augmente indéfiniment, peut-on en conclure que les écarts de ce phénomène suivent la loi normale de répartition ?

Cette question a été résolue rigoureusement par Tchébychef. Nous nous bornerons ici à un problème un peu plus particulier, et plus simple, posé et résolu complètement par Stieltjes, dans un mémoire fondamental ([1]).

Dans ce problème, les intégrales sont prises entre les limites o, ∞. Stieltjes se proposait donc de déterminer une fonction positive $f(x)$, non décroissante, connaissant ses moments

$$c_n = \int_0^{\infty} x^n df(x) \qquad (n = 0, 1, 2, 3, \ldots).$$

Voici comment on peut se rendre compte de la façon dont ce problème se rattache au problème général. Si l'on suppose que l'on se donne les intégrales

$$M_n = \int_{-\infty}^{+\infty} x^n df(x),$$

il est indispensable, pour que ces intégrales aient un sens quel que soit n, que $f(x)$ tende très vite vers la valeur de $f(-\infty)$, supposée nulle, lorsque x tend vers $-\infty$. Dans ces conditions, lorsqu'on voudra déterminer $f(x)$ à l'aide d'un nombre limité de moments M_n, on aura une bonne approximation, facile à évaluer, en se bornant à prendre

$$\int_{-a}^{+\infty} x^n df(x) = M_n,$$

a étant un nombre positif suffisamment grand. Les

$$M_n = \int_{-\infty}^{+\infty} x^n df(x)$$

apparaissent ainsi comme un cas limite des

$$c_n = \int_{-a}^{\infty} x^n df(x).$$

([1]) Cf. *Annales de la Faculté de Toulouse*, t. VIII et IX.

Nous nous proposons de résumer, dans cette etude, les résultats principaux de Stieltjes, sans avoir la prétention de rendre inutile la lecture de son beau mémoire, dans lequel, d'ailleurs, ce problème est la conclusion d'une étude remarquable des fractions continues.

9. Stieltjes arrive à la conclusion que la fonction $f(x)$ est bien déterminée, ou non, suivant qu'une certaine série est divergente ou convergente. Voici comment on peut voir facilement que le problème n'est pas toujours déterminé.

Considérons l'expression

$$d\varphi(x) = e^{-\sqrt[4]{x}} \sin \sqrt[4]{x}\, dx,$$

elle n'est pas constamment positive, mais possède la propriété remarquable d'avoir tous ses moments nuls. Pour vérifier que

$$\gamma_n = \int_0^\infty x^n e^{-\sqrt[4]{x}} \sin \sqrt[4]{x}\, dx = 0,$$

posons $x = z^4$; ce qui donne

$$\int_0^\infty z^{4n+3} e^{-z} \sin z\, dz = 0,$$

ce résultat s'obtient immédiatement en appliquant le théorème fondamental de Cauchy à l'intégrale

$$\int z^{4n+3} e^{-z} (\cos z + i \sin z)\, dz,$$

prise le long du contour OA (C) BO.

Le long de l'arc de cercle (C), de centre O, et de rayon infiniment grand, cette intégrale a une valeur infiniment petite; d'autre part,

$$\int_{BO} = \int_0^\infty z^{4n+3} e^{-z} (-\cos z + i \sin z)\, dz,$$

de sorte que l'intégrale prise le long de tout le contour est égale à $2\gamma_n$, ce qui démontre notre proposition.

On conclut de ce résultat que, si une fonction positive et croissante $f(x)$ admet des moments c_n, la fonction

$$F(x) = f(x) + \lambda \varphi(x),$$

admettra les mêmes moments, quelle que soit la constante λ. Il est clair que $dF(x)$ n'est pas nécessairement positif, pour des valeurs non nulles de λ; par contre, il peut l'être, avec certaines fonctions $f(x)$,

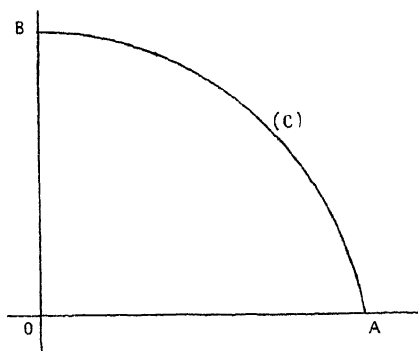



Fig. 26

pour une infinité de valeurs de λ. Il suffit, par exemple, si l'on prend pour $df(x)$ la différentielle $e^{-\sqrt[4]{x}}\,dx$, que λ soit inférieur à 1 en valeur absolue, il y a donc une infinité de fonctions croissantes qui admettent les moments

$$c_n = \int_0^\infty x^n e^{-\sqrt[4]{x}}\,dx$$

10. Ces remarques faites, proposons-nous le problème suivant : « Étant donné un nombre limité de moments, cherchons une fonction non décroissante, en escalier, et définie pour les valeurs positives de x, qui admette ces moments. »

Remarquons d'ailleurs que c'est ce problème qui se pose en pratique, et non la question de la détermination d'une fonction à l'aide d'une infinité de moments.

Désignons par $x_1, x_2, \ldots, x_n$ les points de discontinuité de cette fonction en escalier, les accroissements en ces points étant $a_1, a_2, \ldots, a_n$. Le moment c_p a pour expression

$$c_p = a_1 x_1^p + a_2 x_2^p + \ldots + a_n x_n^p. \tag{1}$$

Pour pouvoir déterminer les $2n$ inconnues positives x_i, a_i ($i = 1, 2, \ldots, n$), il faut se donner $2n$ moments, $c_0, c_1, \ldots, c_{2n-1}$. Il est clair que si l'on se donne, pour moments, des nombres positifs arbi-

traires, les $2n$ équations (1) ($p = 0, 2, \ldots, 2n-1$) n'admettent pas, en général, de solutions positives, ou même réelles. Il peut egalement se présenter des indéterminations

Mais ce qui est remarquable, c'est que « si les c_p sont effectivement les $2n$ premiers moments d'une fonction croissante, les équations (1) admettent des solutions positives ».

C'est ce que montre Stieltjes dans la première partie de son Mémoire, en même temps qu'il donne le moyen d'obtenir ces solutions.

Ce résultat permet de comprendre que l'on puisse considérer une fonction croissante quelconque comme la limite d'une fonction en escalier.

11. Pour nous rendre compte de la méthode de résolution des équations (1), il nous suffira de raisonner sur une valeur déterminée de n, par exemple $n = 3$ Ces équations s'écrivent alors :

$$(2)\quad \begin{cases} c_0 = a_1 + a_2 + a_3, \\ c_1 = a_1 x_1 + a_2 x_2 + a_3 x_3, \\ c_2 = a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2, \\ c_3 = a_1 x_1^3 + a_2 x_2^3 + a_3 x_3^3, \\ c_4 = a_1 x_1^4 + a_2 x_2^4 + a_3 x_3^4, \\ c_5 = a_1 x_1^5 + a_2 x_2^5 + a_3 x_3^5 \end{cases}$$

Elles sont linéaires en a_1, a_2, a_3, ce qui permet d'éliminer facilement ces trois inconnues. D'autre part, ces équations sont symétriques en x_1, x_2, x_3, de sorte que toute fonction symétrique de ces trois variables sera une fonction rationnelle des moments. x_1, x_2, x_3 sont donc les racines d'une équation du troisième degré

$$Q(x) \equiv \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 = 0,$$

dont les coefficients sont des fonctions rationnelles des moments

Ces coefficients sont, d'ailleurs, définis par les trois équations homogènes

$$(3)\quad \begin{cases} \lambda_0 c_0 + \lambda_1 c_1 + \lambda_2 c_2 + \lambda_3 c_3 = 0, \\ \lambda_0 c_1 + \lambda_1 c_2 + \lambda_2 c_3 + \lambda_3 c_4 = 0, \\ \lambda_0 c_2 + \lambda_1 c_3 + \lambda_2 c_4 + \lambda_3 c_5 = 0, \end{cases}$$

que l'on obtient en annulant les expressions $\sum_{i=1}^{3} a_i Q(x_i)$, $\sum_i a_i x_i Q(x_i)$,

$\sum_i a_i x_i^2 Q(x_i)$. On en déduit immédiatement

$$Q(x) \equiv \begin{vmatrix} 1 & x & x^2 & x^3 \\ c_0 & c_1 & c_2 & c_3 \\ c_1 & c_2 & c_3 & c_4 \\ c_2 & c_3 & c_4 & c_5 \end{vmatrix} = 0 \quad (^1).$$

Pour déterminer les masses a_1, a_2, a_3, Stieltjes introduit l'expression symétrique

$$(4) \qquad \frac{P(x)}{Q(x)} = \frac{a_1}{x-x_1} + \frac{a_2}{x-x_2} + \frac{a_3}{x-x_3};$$

$P(x)$ est un polynome du second degré, et l'une quelconque de ces masses est donnée par

$$a_i = \frac{P(x_i)}{Q'(x_i)}.$$

Or

$$\begin{aligned} P(x) &= a_1 \frac{Q(x)}{x-x_1} + a_2 \frac{Q(x)}{x-x_2} + a_3 \frac{Q(x)}{x-x_3} \\ &= a_1[\lambda_3 x^2 + (\lambda_3 x_1 + \lambda_2)x + (\lambda_3 x_1^2 + \lambda_2 x_1 + \lambda_1)] \\ &\quad + a_2[\lambda_3 x^2 + (\lambda_3 x_2 + \lambda_2)x + (\lambda_3 x_2^2 + \lambda_2 x_2 + \lambda_1)] \\ &\quad + a_3[\lambda_3 x^2 + (\lambda_3 x_3 + \lambda_2)x + (\lambda_3 x_3^2 + \lambda_2 x_3 + \lambda_1)], \end{aligned}$$

d'où l'on déduit

$$P(x) = c_0 \lambda_3 x^2 + (\lambda_3 c_1 + \lambda_2 c_0)x + (\lambda_3 c_2 + \lambda_2 c_1 + \lambda_1 c_0),$$

ou encore

$$P(x) = \begin{vmatrix} 0 & c_0 & c_0 x + c_1 & c_0 x^2 + c_1 x + c_2 \\ c_0 & c_1 & c_2 & c_3 \\ c_1 & c_2 & c_3 & c_4 \\ c_2 & c_3 & c_4 & c_5 \end{vmatrix} = 0.$$

Pour démontrer que les x_i et les a_i sont positifs, nous serons ainsi conduits à vérifier que le polynome $Q(x) = 0$ a ses trois racines positives et séparées par les deux racines de $P(x) = 0$.

12. Ceci posé, plaçons-nous dans le cas, qui nous intéresse ici,

---

(1) Le polynome, que nous désignerons par $Q(x)$, serait désigné par Stieltjes, avec la notation de son Mémoire, par $Q_6(-x)$, à un facteur constant près. La même remarque peut être faite relativement à la notation du polynome $P(x)$ dont il est question ensuite.

où $c_0, c_1, \ldots, c_5$ sont effectivement les six premiers moments d'une fonction croissante $f(x) \cdot c_n = \int_0^\infty x^n df(x)$.

Dans ces conditions, les équations (3) prennent la forme

$$(3') \quad \begin{cases} \int_0^\infty (\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3)\, df(x) = 0, \\ \int_0^\infty (\lambda_0 x + \lambda_1 x^2 + \lambda_2 x^3 + \lambda_3 x^4)\, df(x) = 0, \\ \int_0^\infty (\lambda_0 x^2 + \lambda_1 x^3 + \lambda_2 x^4 + \lambda_3 x^5)\, df(x) = 0, \end{cases}$$

ou encore

$$(3'') \quad \begin{cases} \int_0^\infty Q(x)\, df(x) = 0, \\ \int_0^\infty Q(x) x\, df(x) = 0, \\ \int_0^\infty Q(x) x^2 df(x) = 0. \end{cases}$$

Le polynome du troisième degré $Q(x)$ est donc tel que l'intégrale $\int_0^\infty Q(x)F(x)df(x)$ est nulle lorsque $F(x)$ est un polynome arbitraire du second degré en x.

D'une manière générale, si l'on se donnait $2n$ moments, on prendrait n escaliers, $Q(x)$ serait de degré n, et $F(x)$ serait un polynome arbitraire de degré $(n-1)$.

La fonction $f(x)$ étant donnée, on peut attribuer à n la succession des valeurs entières, et former ainsi une suite de polynomes $Q_k(x)$ $(k = 1, 2, 3, \ldots)$, de degré k, tels que l'on ait

$$\int_0^\infty Q_k(x)\, F_{k-1}(x)\, df(x) = 0,$$

pour tout polynome $F_{k-1}(x)$ de degré $k-1$ au plus.

En particulier, deux polynomes distincts de cette suite satisfont à l'identité

$$\int_0^\infty Q_k(x) Q_h(x)\, df(x) = 0 \quad (k \neq h);$$

on rappelle cette propriété en disant que ces polynomes constituent une suite orthogonale relativement à la fonction $f(x)$.

13. Dans les raisonnements précédents, nous avons supposé implicitement que le polynome $Q_k(x)$ est effectivement de degré k. Ceci suppose que les déterminants

$$\begin{vmatrix} c_0 & c_1 & c_2 & & c_{k-1} \\ c_1 & c_2 & c_3 & \cdots & c_k \\ \cdots & \cdots & & \cdots & \\ c_{k-1} & c_k & c_{k+1} & \cdot & c_{2k-2} \end{vmatrix}$$

ne peuvent être nuls; or, il est facile de voir qu'ils sont essentiellement positifs. Vérifions-le, dans l'exemple que nous avons choisi, où $k = 3$. On peut écrire, en effet,

$$\lambda_3 = \begin{vmatrix} \int_0^\infty df(x) & \int_0^\infty x\, df(x) & \int_0^\infty x^2 df(x) \\ \int_0^\infty y\, df(y) & \int_0^\infty y^2 df(y) & \int_0^\infty y^3 df(y) \\ \int_0^\infty z^2 df(z) & \int_0^\infty z^3 df(z) & \int_0^\infty z^4 df(z) \end{vmatrix},$$

λ_3 apparaît alors comme une somme d'intégrales triples étendues au premier trièdre de l'espace (x, y, z); ou encore

$$\lambda_3 = \int_0^\infty \int_0^\infty \int_0^\infty \begin{vmatrix} df(x) & x\, df(x) & x^2 df(x) \\ y\, df(y) & y^2 df(y) & y^3 df(y) \\ z^2 df(z) & z^3 df(z) & z^4 df(z) \end{vmatrix}$$

$$= \int_0^\infty \int_0^\infty \int_0^\infty \begin{vmatrix} 1 & x & x^2 \\ y & y^2 & y^3 \\ z^2 & z^3 & z^4 \end{vmatrix} df(x)\, df(y)\, df(z).$$

Si nous posons

$$\delta = \begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{vmatrix},$$

λ_3 prend la forme

$$\lambda_3 = \int_0^\infty \int_0^\infty \int_0^\infty y z^2 \delta\, df(x)\, df(y)\, df(z);$$

or, en permutant les rôles de y et z, on aurait obtenu

$$\lambda_3 = \int_0^\infty \int_0^\infty \int_0^\infty (-y^2 z)\delta .\, df(x)\, df(y)\, df(z);$$

on peut permuter x, y, z, deux à deux, de 3! façons différentes, et ajouter entre elles les expressions obtenues, ce qui donne immédiatement

$$3!\lambda_3 = \int_0^\infty \int_0^\infty \int_0^\infty \delta^2 \, df(x) \, df(y) \, df(z),$$

expression essentiellement positive, non nulle, si $f(x)$ n'est pas en escalier.

Remarque. — Cette conclusion peut être en défaut si la fonction $f(x)$ est elle-même une fonction en escalier. Si p est le nombre de ses escaliers, l'intégrale

$$n!\lambda_n = \int_0^\infty \int_0^\infty \cdots \int_0^\infty \begin{vmatrix} 1 & x_1 & x_1^2 & & x_1^{n-1} \\ 1 & x_2 & x_2^2 & & x_2^{n-1} \\ & & & & \\ 1 & x_n & x_n^2 & & x_n^{n-1} \end{vmatrix} df(x_1) \, df(x_2) \, . \, df(x_n)$$

est évidemment nulle dès que n surpasse p. Dans ce cas, $Q_n(x)$ reste de degré p lorsque n est supérieur ou égal à p. Ce polynome est alors confondu avec celui qui fournit les abscisses x_i des escaliers de la fonction donnée $f(x)$, et les a_i sont les masses $df(x_i)$ elles-mêmes.

Le cas où $f(x)$ est en escalier est donc un cas exceptionnel sans intérêt et que nous écarterons dans la suite Cette remarque s'appliquerait aussi bien si l'intervalle $(0, \infty)$ était remplacé par un intervalle fini (a, b), et là encore nous écarterons les fonctions $f(x)$ en escalier.

14. Voici une conséquence immédiate et intéressante du résultat que nous avons démontré dans le paragraphe précédent.

Par construction, $Q_n(x)$ a les mêmes $2n$ premiers moments que la fonction donnée $f(x)$. Il n'en est plus de même pour son $(2n+1)^{\text{ième}}$ moment; il est facile de voir, en effet, qu'il est inférieur à celui de $f(x)$. Désignons-le par γ_{2n}, et soit

$$Q_n(x) = \lambda_0 + \lambda_1 x + \ldots + \lambda_n x^n,$$

il vient alors

$$\begin{cases} \lambda_0 c_0 + \lambda_1 c_1 + \ldots + \lambda_n c_n = 0, \\ \lambda_0 c_1 + \lambda_1 c_2 + \ldots + \lambda_n c_{n+1} = 0, \\ \ldots \ldots \ldots \ldots \ldots \ldots \ldots, \\ \lambda_0 c_{n-1} + \lambda_1 c_n + \ldots + \lambda_n c_{2n-1} = 0, \\ \lambda_0 c_n + \lambda_1 c_{n+1} + \ldots + \lambda_n \gamma_{2n} = 0, \end{cases}$$

ces équations, homogènes en $\lambda_0, \lambda_1, \ldots, \lambda_n$, exigent que le déterminant de leurs coefficients soit nul, ce qui définit γ_{2n} par l'équation du premier degré

$$\begin{vmatrix} c_0 & c_1 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n-1} & c_n & \cdots & c_{2n-1} \\ c_n & c_{n+1} & \cdots & \gamma_{2n} \end{vmatrix} = 0$$

Or, le même déterminant, où γ_{2n} est remplacé par c_{2n}, est positif, ce qui entraîne l'inégalité

$$\gamma_{2n} \begin{vmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n-1} & c_n & \cdots & c_{2n-2} \end{vmatrix} \leqq c_{2n} \begin{vmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n-1} & c_n & \cdots & c_{2n-2} \end{vmatrix},$$

et, par suite,

$$\gamma_{2n} \leqq c_{2n},$$

puisque le coefficient commun est positif.

Remarquons encore que si $f(x)$ était une fonction à p escaliers, cette inégalité devrait être remplacée par l'égalité pour γ_{2p}, ainsi que pour les moments d'ordres supérieurs

15. Étude des polynomes $Q_k(x)$. — La propriété que nous avons démontrée, de la suite des $Q_k(x)$, nous fait prévoir que leur étude sera analogue à celle des polynomes de Legendre. La méthode employée dans la théorie de ces derniers pourra nous servir, ici, à démontrer les propriétés des zéros des $Q_k(x)$.

Nous considérerons, d'une manière plus générale, des polynomes $Q_n(x)$, de degré n, tels que, $f(x)$ étant une certaine fonction, croissante dans l'intervalle (a, b), l'égalité

$$\int_a^b Q_n(x) F(x)\, df(x) = 0$$

soit satisfaite pour tout polynome $F(x)$ de degré inférieur à n.

Montrons que, dans ces conditions, $Q_n(x) = 0$ a ses n racines réelles, distinctes, et comprises entre a et b.

Tout d'abord, $\int_a^b Q_n(x)\, df(x) = 0$ exige que $Q_n(x)$ n'ait pas un

signe constant entre a et b. Soient $x_1, x_2, \ldots, x_k$ $(k \leqq n)$ les racines de $Q_n(x) = 0$, comprises entre a et b, une racine multiple étant comptée un nombre de fois égal à son degré de multiplicité. On peut poser

$$Q_n(x) = (x - x_1)(x - x_2)\ldots(x - x_k)\,q_n(x),$$

avec

$$a < x_1 \leqq x_2 \leqq \ldots \leqq x_k < b$$

et $q_n(x)$ ayant un signe constant dans l'intervalle (a, b). Si k était inférieur à n il suffirait de prendre pour $F(x)$ le polynome $(x - x_1)(x - x_2)\ldots(x - x_k)$ pour avoir

$$\int_a^b (x - x_1)\ldots(x - x_k)\,Q_n(x)\,df(x)$$
$$= \int_a^b (x - x_1)^2 \ldots (x - x_k)^2\,q_n(x)\,df(x) = 0,$$

ce qui est impossible.

Le même raisonnement permet de voir qu'il ne peut y avoir aucune racine multiple. En effet, $Q_n(x)$ ne peut pas avoir que des racines d'ordre pair, puisqu'il n'a pas un signe constant dans (a, b); désignons alors par $x_{\alpha_1}, x_{\alpha_2}, \ldots, x_{\alpha_h}$ toutes les racines distinctes, d'ordre de multiplicité impair; h serait nécessairement inférieur à n si $Q(x)$ avait des racines multiples, ce qui permettrait de prendre pour $F(x)$ le polynome $(x - x_{\alpha_1})(x - x_{\alpha_2})\ldots(x - x_{\alpha_h})$ et d'arriver à une contradiction, en écrivant que

$$\int_a^b Q_n(x)(x - x_{\alpha_1})(x - x_{\alpha_2})\ldots(x - x_{\alpha_n})\,df(x) = 0$$

16. Étudions maintenant la situation respective des zéros de deux polynomes $Q_n(x)$ consécutifs. Remarquons tout d'abord que $Q_n(x)$ étant effectivement de degré n, nous pouvons supposer que son coefficient de x^n est réduit à l'unité; c'est ce que nous ferons dans ce qui suit.

Nous allons voir que les zéros de $Q_n(x)$ sont séparés par les $(n - 1)$ zéros de $Q_{n-1}(x)$. Pour cela, effectuons la division de ces deux polynomes entre eux; il vient

$$Q_n(x) = (x - \alpha_{n-1})\,Q_{n-1}(x) + R(x),$$

$R(x)$ étant de degré $(n - 2)$ au plus, et α_{n-1} étant une certaine

constante. Si k est un nombre entier inférieur à $n-2$, l'identité

$$\begin{aligned} 0 &= \int_a^b x^k Q_n(x)\,df(x) \\ &= \int_a^b (x-\alpha_{n-1}) x^k Q_{n-1}(x)\,df(x) + \int_a^b x^k R(x)\,df(x) \end{aligned}$$

se réduit à

$$\int_a^b x^k R(x)\,df(x) = 0 \qquad (k = 0, 1, 2, \ldots, n-3).$$

Ces dernières relations sont caractéristiques de $Q_{n-2}(x)$ et des polynomes qui lui sont proportionnels; on peut donc poser

$$R(x) = -\lambda_{n-1} Q_{n-2}(x).$$

La division effectuée plus haut prend alors la forme

$$Q_n(x) = (x-\alpha_{n-1}) Q_{n-1}(x) - \lambda_{n-1} Q_{n-2}(x), \tag{5}$$

et nous allons déterminer les deux constantes α_{n-1} et λ_{n-1}.

Pour éliminer λ_{n-1}, multiplions les deux membres de la relation (5) par $Q_{n-1}(x)\,df(x)$, et intégrons entre a et b. Il vient

$$0 = \int_a^b (x-\alpha_{n-1}) Q_{n-1}(x)^2\,df(x),$$

ce qui donne la valeur de α_{n-1} :

$$\alpha_{n-1} = \frac{\displaystyle\int_a^b x Q_{n-1}(x)^2\,df(x)}{\displaystyle\int_a^b Q_{n-1}(x)^2\,df(x)};$$

ceci nous montre, d'ailleurs, que α_{n-1} est compris entre a et b. Pour calculer λ_{n-1}, multiplions les deux membres de (5) par $Q_{n-2}(x)\,df(x)$; l'intégration entre a et b donne alors

$$0 = \int_a^b x Q_{n-1}(x) Q_{n-2}(x)\,df(x) - \lambda_{n-1} \int_a^b Q_{n-2}(x)^2\,df(x).$$

Or, $x Q_{n-2}$ diffère de Q_{n-1} par un polynome de degré $n-2$ au plus; on a donc l'identité

$$\int_a^b x Q_{n-2} Q_{n-1}\,df(x) = \int_a^b Q_{n-1}^2\,df(x).$$

ce qui donne enfin

$$\lambda_{n-1} = \frac{\int_a^b Q_{n-1}(x)^2 \, df(x)}{\int_a^b Q_{n-2}(x)^2 \, df(x)},$$

et l'on voit que les λ_{n-1} sont essentiellement positifs.

Ces résultats obtenus, considérons la succession des polynomes

$$\left\{\begin{array}{ll} Q_0(x) = 1, & \\ Q_1(x) = x - \alpha_0 & (a < \alpha_0 < b), \\ Q_2(x) = (x - \alpha_1) Q_1(x) - \lambda_1 & (a < \alpha_1 < b, \quad \lambda_1 > 0), \\ Q_3(x) = (x - \alpha_2) Q_2(x) - \lambda_2 Q_1(x) & (a < \alpha_2 < b, \quad \lambda_2 > 0), \\ \dots & \dots \end{array}\right.$$

$Q_1(x)$ s'annule lorsque x prend la valeur α_0, comprise entre a et b, mais, pour cette même valeur, $Q_2(\alpha_0) = -\lambda_1 < 0$ et $Q_2(x)$ étant positif à l'infini, α_0 sépare les deux zéros x_1, x_2 de $Q_2(x)$.

On sait déjà que x_1 et x_2 sont compris entre a et b, ce qui permet d'effectuer le classement

$$a < x_1 < \alpha_0 < x_2 < b.$$

Formons maintenant le tableau des valeurs de $Q_3(x)$ pour la succession de valeurs de x : $-\infty$, x_1, x_2, $+\infty$. On obtient le tableau suivant

x.	$-\infty$	x_1	x_2.	$+\infty$
$Q_3(x)$	$-$	$-\lambda_2 Q_1(x_1)$ $+$	$-\lambda_2 Q_1(x_2)$ $-$	$+$

qui nous montre immédiatement que les trois zéros x'_1, x'_2, x'_3 de ce polynome sont séparés par x_1, x_2, ou, plus précisément, se classent de la façon suivante dans l'intervalle (a, b)

$$a < x'_1 < x_1 < x'_2 < x_2 < x'_3 < b.$$

Il est inutile de pousser plus loin cette étude pour se rendre compte de la généralité de ce raisonnement, qui nous montrerait, d'une façon générale, que les racines de $Q_n(x) = 0$ sont séparées par celles de $Q_{n-1}(x) = 0$.

17. Étude des masses a_i. — Occupons-nous maintenant du signe

des sauts $a_1, a_2, \ldots, a_n$, aux points d'abscisses $x_1, x_2, \ldots, x_n$ et montrons qu'ils sont tous positifs. Nous continuons à supposer que les $c_0, c_1, c_2, \ldots$ sont effectivement les moments d'une fonction croissante $f(x)$ définie entre a et b.

Proposons-nous d'évaluer l'intégrale $\int_a^b G(x)\,df(x)$, dans laquelle $G(x)$ est un polynome entier, de degré $2n - 1$ au plus. Cette évaluation conduit à une expression très simple, si l'on introduit le polynome $Q_n(x)$ dont nous désignerons toujours les zéros par $x_1, x_2, \ldots, x_n$.

Si nous divisons $G(x)$ par $Q_n(x)$, le quotient $g(x)$ et le reste $R(x)$ sont de degré $(n-1)$ au plus, et l'on a

$$G(x) = Q_n(x)\,g(x) + R(x)$$

Il en résulte l'identité

$$\int_a^b G(x)\,df(x) = \int_a^b R(x)\,df(x),$$

qui nous montre que l'expression cherchée dépend de n constantes arbitraires au plus; car, le polynome $R(x)$ étant défini dès que l'on connaît $R(x_1), R(x_2), \ldots, R(x_n)$, nous pouvons exprimer notre intégrale en fonction de ces n valeurs.

D'ailleurs, d'après la formule d'interpolation de Lagrange, on peut mettre $R(x)$ sous la forme

$$R(x) = \sum_{i=1}^{n} R(x_i)\,\frac{Q_n(x)}{Q'_n(x_i)\,(x - x_i)},$$

ce qui nous permet d'écrire

$$\int_a^b G(x)\,df(x) = \sum_{i=1}^{n} R(x_i) \int_a^b \frac{Q_n(x)}{(x - x_i)\,Q'_n(x_i)}\,df(x),$$

ou encore

$$\int_a^b G(x)\,df(x) = \sum_{i=1}^{n} A_i\,R(x_i),$$

en posant

$$(6) \qquad A_i = \int_a^b \frac{Q_n(x)}{(x - x_i)\,Q'_n(x_i)}\,df(x);$$

les constantes A_i sont indépendantes de $G(x)$; elles ne dépendent que de $f(x)$ et de n.

Si nous remarquons que $G(x_i) = R(x_i)$, nous pouvons écrire finalement

$$(7) \qquad \int_a^b G(x)\,df(x) = A_1\,G(x_1) + A_2\,G(x_2) + \ldots + A_n\,G(x_n),$$

sous la seule condition que le degré du polynome $G(x)$ ne surpasse pas $2n-1$.

En particulier, si l'on prend pour $G(x)$ le monome x^m ($m = 0, 1, 2, \ldots, 2n-1$), la relation (7) donne la suite de relations

$$\int_a^b x^m\,df(x) = A_1 x_1^m + A_2 x_2^m + \ldots + A_n x_n^m,$$

grâce auxquelles les A_i nous apparaissent comme n'étant rien autre que les sauts a_i de la fonction, à n escaliers, qui admet les mêmes $2n$ premiers moments que $f(x)$, et les formules (6) nous font connaître l'expression de ces masses a_i.

Pour vérifier que les a_i sont positifs, nous allons en donner une autre expression, que nous déduirons de l'application de l'identité (7) aux $(n-1)$ polynomes

$$Q_{n-1}(x), \quad x\,Q_{n-1}(x), \quad \ldots, \quad x^{n-2}Q_{n-1}(x).$$

D'ailleurs, nous savons que $\int_a^b x^k Q_{n-1}(x)\,df(x)$ est nul pour les valeurs entières de k inférieures ou égales à $(n-2)$, de sorte que nous obtenons ainsi $(n-1)$ équations en $a_1, a_2, \ldots, a_n$,

$$(8) \quad \left\{ \begin{array}{l} a_1\,Q_{n-1}(x_1) + a_2\,Q_{n-1}(x_2) + \ldots + a_n\,Q_{n-1}(x_n) = 0, \\ a_1 x_1\,Q_{n-1}(x_1) + a_2 x_2\,Q_{n-1}(x_2) + \ldots + a_n x_n\,Q_{n-1}(x_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ a_1 x_1^{n-2}\,Q_{n-1}(x_1) + a_2 x_2^{n-2}\,Q_{n-1}(x_2) + \ldots + a_n x_n^{n-2}\,Q_{n-1}(x_n) = 0; \end{array} \right.$$

ces équations homogènes définissent les a_i à un facteur près, car les coefficients de ces inconnues sont tous différents de zéro, et, d'autre part, les $x_1, x_2, \ldots, x_n$ étant distincts, les déterminants de ces coefficients ne sont pas nuls.

Pour résoudre les équations (8) en a_1 et a_2, introduisons les coefficients $\lambda_0, \lambda_1, \ldots, \lambda_{n-2}$ du quotient de $Q_n(x)$ par $(x-x_1)(x-x_2)$;

soit

$$\frac{Q_n(x)}{(x-x_1)(x-x_2)} = \lambda_0 + \lambda_1 x + \ldots + \lambda_{n-3} x^{n-3} + \lambda_{n-2} x^{n-2},$$

le coefficient λ_{n-2} étant égal à 1, d'après l'hypothèse faite sur le coefficient de x^n dans $Q_n(x)$. Multiplions alors les équations (8) par λ_0, λ_1, ..., λ_{n-2} et additionnons-les; il vient

$$a_1 Q_{n-1}(x_1)\frac{Q'_n(x_1)}{x_1-x_2} + a_2 Q_{n-1}(x_2)\frac{Q'_n(x_2)}{x_2-x_1} = 0,$$

c'est-à-dire

$$a_1 Q'_n(x_1) Q_{n-1}(x_1) = a_2 Q'_n(x_2) Q_{n-1}(x_2);$$

et, par suite, en désignant par Λ cette valeur commune, on aura d'une manière générale

$$(9) \qquad a_i Q'_n(x_i) Q_{n-1}(x_i) = \Lambda \qquad (i = 1, 2, \ldots, n).$$

Pour calculer Λ, appliquons la formule (7) au polynome particulier $G(x) = Q'_n(x) Q_{n-1}(x)$. Nous obtenons

$$\int_a^b Q'_n(x) Q_{n-1}(x)\, df(x) = n\Lambda.$$

Le premier terme de $Q_n(x)$ étant x^n, $Q'_n(x)$ ne diffère de $nQ_{n-1}(x)$ que par un polynome de degré $(n-2)$ au plus, ce qui nous donne enfin

$$(10) \qquad \Lambda = \int_a^b Q_{n-1}(x)^2\, df(x),$$

dont la valeur est essentiellement positive.

Pour montrer que les a_i sont positifs, il reste à vérifier que leurs coefficients $Q'_n(x_i) Q_{n-1}(x_i)$ sont positifs. Or, ceci résulte de ce que les zéros de $Q'_n(x)$, aussi bien que ceux de $Q_{n-1}(x)$, séparent les zéros de $Q_n(x)$, de sorte que les $Q'_n(x_i) Q_{n-1}(x_i)$ sont tous de même signe; ce signe est d'ailleurs le signe +, puisque, en particulier, $Q'_n(x_n) Q_{n-1}(x_n)$ est positif.

« En résumé, la fonction $f(x)$ est remplaçable, dans une certaine mesure, par une fonction en escalier; on peut dire encore qu'une répartition quelconque de masse, sur une demi-droite, est remplaçable par un nombre fini de masses isolées; ce nombre de masses devant être d'autant plus grand que les problèmes font intervenir des polynomes de plus haut degré. »

18. Les formules précédentes peuvent être remplacées par une seule, en se plaçant au point de vue formel, ceci signifie que les développements que nous allons écrire ne sont pas nécessairement convergents, mais que l'égalité que nous écrirons de deux séries entières équivaudra à leur identité terme à terme, quelle que soit la nature de ces séries.

Dans ces conditions, on peut remplacer les formules de définition $c_k = \int_a^b x^k df(x)$ $(k = 0, 1, 2, \ldots)$ par l'identité formelle

$$\sum_{k=0}^{\infty} c_k u^k = \int_a^b \sum_{k=0}^{\infty} u^k x^k \, df(x) = \int_a^b \frac{df(x)}{1 - ux}.$$

Les développements ne sont convergents que si u est assez petit en valeur absolue, mais, par hypothèse, l'identification terme à terme est possible quelle que soit la valeur de u.

Lorsqu'on remplace $f(x)$ par la fonction en escalier définie par le polynome $Q_n(x)$ et les sauts $a_1, a_2, \ldots, a_n$, l'intégrale $\int_a^b \frac{df(x)}{1 - ux}$ est remplacée par la somme

$$\begin{aligned}
&\frac{a_1}{1 - ux_1} + \frac{a_2}{1 - ux_2} + \ldots + \frac{a_n}{1 - ux_n} \\
&= \sum_{k=0}^{\infty} u^k (a_1 x_1^k + a_2 x_2^k + \ldots + a_n x_n^k) \\
&= \sum_{k=0}^{2n-1} c_k u^k + \sum_{k=2n}^{\infty} u^k (a_1 x_1^k + a_2 x_2^k + \ldots + a_n x_n^k).
\end{aligned}$$

$Q_n(x)$ peut être considéré comme représentant la fonction de n escaliers $\varphi(x)$ pour laquelle le développement de l'intégrale $\int_a^b \frac{d\varphi(x)}{1 - ux}$ a les mêmes $2n$ premiers termes que le développement relatif à la fonction donnée $f(x)$.

La forme précédente est, à la notation près, la forme adoptée par Stieltjes. Mais on peut adopter beaucoup d'autres développements formels.

On pourrait considérer, par exemple, le développement

$$\sum_{k=0} \frac{c_k u^k}{k!} = \int_a^b \sum_{k=0}^{\infty} \frac{u^k x^k}{k!} df(x) = \int_a^b e^{ux} \, df(x).$$

La même intégrale, calculée avec $Q_n(x)$, donne alors la somme

$$\sum_{i=1}^{n} a_i e^{u x_i} = \sum_{k=0}^{\infty} \frac{u^k}{k!}(a_1 x_1^k + a_2 x_2^k + \dots + a_n x_n^k)$$
$$= \sum_{k=0}^{2n-1} \frac{c_k u^k}{k!} + \sum_{k=2n}^{\infty} \frac{u^k}{k!}(a_1 x_1^k + a_2 x_2^k + \dots + a_n x_n^k)$$

La fonction à n escaliers définie par $Q_n(x)$ correspond encore à l'identité des $2n$ premiers termes, du développement de l'intégrale $\int_a^b e^{ux} d\varphi(x)$ relative à cette fonction, et du développement de $\int_a^b e^{ux} df(x)$.

19. Occupons-nous de la convergence du développement de Stieltjes $\sum_{k=0}^{\infty} c_k u^k$; elle dépend de l'ordre de croissance des c_n. Nous allons l'étudier dans le cas, qui nous intéresse plus particulièrement, où l'intervalle (a, b) est l'intervalle $(0, \infty)$.

En supposant toujours que les c_k sont effectivement les moments d'une fonction $f(x)$, tous les déterminants

$$\begin{vmatrix} c_p & c_{p+1} & \dots & c_{p+n} \\ c_{p+1} & c_{p+2} & \dots & c_{p+n+1} \\ \dots & \dots & \dots & \dots \\ c_{p+n} & c_{p+n+1} & \dots & c_{p+2n} \end{vmatrix}$$

sont positifs.

En effet, en remplaçant c_k par son expression $\int_0^{\infty} x^k df(x)$, on peut écrire ce déterminant, d'après un calcul déjà fait dans le cas particulier $p = 0$,

$$\int\int\dots\int \begin{vmatrix} x^p & x^{p+1} & \dots & x^{p+n} \\ y^{p+1} & y^{p+2} & \dots & y^{p+n+1} \\ \dots & \dots & \dots & \dots \\ t^{p+n} & t^{p+n+1} & \dots & t^{p+2n} \end{vmatrix} df(x)\, df(y) \dots df(t),$$

ou, en posant

$$\Delta = \begin{vmatrix} 1 & x & \dots & x^n \\ 1 & y & \dots & y^n \\ \dots & \dots & \dots & \dots \\ 1 & t & \dots & t^n \end{vmatrix}$$

$$\int\int \dots \int x^p y^{p+1} \dots t^{p+n} \Delta \, df(x) \, df(y) \dots df(t).$$

Il suffit maintenant de permuter de toutes les façons possibles les n variables $x, y, \dots, t$, ce qui fournit $n!$ expressions différentes de notre déterminant, et d'ajouter toutes ces expressions entre elles. On obtient ainsi, pour leur valeur commune,

$$\frac{1}{n!} \int\int \dots \int x^p y^p \dots t^p \Delta^2 \, df(x) \, df(y) \dots df(t)$$

expression qui est bien positive, puisque $x, y, \dots, t$ varient de 0 à ∞.

Il résulte, d'ailleurs, de cette même expression que, si p est un nombre pair, le déterminant en question est positif, pour l'intervalle de variation (a, b) le plus général. Si p est impair, ce déterminant est encore positif pour tout intervalle (a, b) compris dans l'intervalle $(0, \infty)$: $0 \leqq a < b$.

En particulier, pour $n = 2$, nous avons

$$c_p c_{p+2} - c_{p+1}^2 > 0,$$

et, par suite,

$$\frac{c_{p+2}}{c_{p+1}} > \frac{c_{p+1}}{c_p}; \tag{11}$$

donc le rapport $\frac{c_{p+1}}{c_p}$ de deux moments consécutifs croît constamment avec p. Ce rapport admet donc une limite L, finie ou non.

Si cette limite L est infinie, la série $\sum_{k=0}^{\infty} c_k u^k$ est divergente quel que soit u; dans le cas contraire, cette série a un rayon de convergence égal à $\frac{1}{L}$.

On peut encore déduire de ceci que $\sqrt[p]{c_p}$ croît constamment avec p.

Posons $\gamma_p = \sqrt[p]{c_p}$. L'inégalité (11) s'écrit alors

$$\frac{\gamma_{p+1}^{p+1}}{\gamma_p^p} > \frac{\gamma_p^p}{\gamma_{p-1}^{p-1}}$$

ou

$$\left(\frac{\gamma_{p+1}}{\gamma_p}\right)^{p+1} > \left(\frac{\gamma_p}{\gamma_{p-1}}\right)^{p-1}.$$

Donc, si $\gamma_p > \gamma_{p-1}$, $\left(\frac{\gamma_p}{\gamma_{p-1}}\right)^{p-1}$ est supérieur à 1, et il en est de même, *a fortiori* de $\left(\frac{\gamma_{p+1}}{\gamma_p}\right)^{p+1}$, ce qui entraîne $\gamma_{p+1} > \gamma_p$. Il suffit donc de vérifier que l'on a $\gamma_2 > \gamma_1$; or, si c_0 est égal à l'unité, ce qu'on peut toujours supposer, l'inégalité

$$c_0 c_2 > c_1^2$$

est équivalente à $\gamma_2 > \gamma_1$; ce qui démontre que γ_p est une fonction croissante de p

Tous ces calculs permettent encore d'obtenir quelques renseignements sur la variation de la plus grande racine x_n du polynome $Q_n(x)$, lorsque n varie. Nous avons, en effet, pour $p \leqq 2n-2$,

$$\frac{c_{p+1}}{c_p} = \frac{a_1 x_1^{p+1} + a_2 x_2^{p+1} + \ldots + a_n x_n^{p+1}}{a_1 x_1^p + a_2 x_2^p + \ldots + a_n x_n^p};$$

donc $\frac{c_{2n-1}}{c_{2n-2}}$ est compris entre le plus petit et le plus grand des rapports $\frac{a_i x_i^{2n-1}}{a_i x_i^{2n-2}} = x_i$, c'est-à-dire entre x_1 et x_n. Donc, si $\frac{c_{2n-1}}{c_{2n-2}}$ augmente indéfiniment avec n, il en est nécessairement de même de x_n

La réciproque de cette proposition est d'ailleurs exacte, mais nous ne la démontrerons pas ([1]).

20. Reprenons la formule générale

$$(7) \qquad \int_b^a G(x)\,df(x) = a_1 G(x_1) + a_2 G(x_2) + \ldots + a_n G(x_n),$$

dans laquelle $G(x)$ est un polynome arbitraire de degré $2n-1$ au plus;

([1]) Pour la démonstration de cette réciproque, *cf.* Stieltjes, Chapitre II du Mémoire déjà cité

nous savons également que les x_i sont intérieurs à l'intervalle (a, b).

$$0 < a < x_1 < x_2 < \ldots < x_n < b$$

Nous allons en déduire des résultats très importants en particularisant le polynome $G(x)$. Choisissons pour $G(x)$ un polynome $T(x)$, de degré $2n-2$, vérifiant les propriétés suivantes

$$(12)\qquad \begin{cases} T(x) > 1 & \text{pour}\quad a < x < x_k,\\ T(x) > 0 & \text{pour}\quad x_k < x < b.\end{cases}$$

x_k désignant une des racines de $Q_n(x) = 0$

Pour réaliser ces inégalités, nous définirons $T(x)$ par des conditions d'égalité, quitte à vérifier ensuite que les inégalités en question sont satisfaites. Imposons à ce polynome les n conditions

$$(13)\qquad \begin{cases} T(x_1) \;= T(x_2) \;= \ldots = T(x_k) = 1,\\ T(x_{k+1}) = T(x_{k+2}) = \ldots = T(x_n) = 0\end{cases}$$

il reste encore $n-1$ coefficients arbitraires, que nous déterminerons par les $(n-1)$ conditions supplémentaires

$$(14)\qquad \begin{cases} T'(x_1) \;= T'(x_2) \;= \ldots = T'(x_{k-1}) = 0,\\ T'(x_{k+1}) = T'(x_{k+2}) = \ldots = T'(x_n) \;= 0.\end{cases}$$

Les relations (13) et (14) étant supposées vérifiées, la dérivée $T'(x)$ s'annule nécessairement dans chacun des intervalles

$$(x_1, x_2),\quad (x_2, x_3),\quad \ldots,\quad (x_{k-1}, x_k),\quad (x_{k+1}, x_{k+2}),\quad \ldots,\quad (x_{n-1}, x_n),$$

autrement dit, dans tous les intervalles déterminés par les zéros de $Q_n(x)$, sauf l'intervalle (x_k, x_{k+1}). En désignant par $\xi_1, \xi_2, \ldots, \xi_{n-1}$ les racines de $T'(x) = 0$ situées dans ces $(n-2)$ intervalles, les $2n-3$ zéros de $T'(x)$ sont tous les nombres de la suite

$$x_1 < \xi_1 < x_2 < \xi_2 < x_3 < \ldots < x_{k-1} < \xi_{k-1} < x_{k+1} < \xi_{k+1} < x_{k+2} < \ldots \\ < x_{n-1} < \xi_{n-1} < x_n,$$

et situées, les $2k-2$ premiers, entre a et x_k, et les $2n-2k-1$ derniers entre x_{k+1} et b.

Il ne peut donc y en avoir aucun entre x_k et x_{k+1}, et $T(x_k)$ étant égal à 1, donc supérieur à $T(x_{k+1})$ qui est nul, $T'(x)$ est constam-

ment négatif entre ξ_{k-1} et x_{k+1}. Ces résultats permettent de former le tableau de variation de $T'(x)$ et de $T(x)$.

x	a		x_1		ξ_1		x_2	
$T'(x)$		—	0	+	0	—	0	
$T(x)$		↘	1	↗		↘	1	. .

x	x_{k-1}		ξ_{k-1}		x_k		x_{k+1}		ξ_{k+1}		x_{k+2}	...
$T'(x)$	0	+	0	—		—	0	+	0	—	0	.
$T(x)$	1	↗		↘	1	↘	0	↗		↘	0	. .

x	x_{n-1}		ξ_{n-1}		x_n		b
$T'(x)$	0	+	0	—	0	+	
$T(x)$	0	↗		↘	0	↗	

La variation de $T(x)$ aura donc l'allure de la courbe représentée sur la figure ci-dessous, qui met immédiatement en évidence la vérification des inégalités (12)

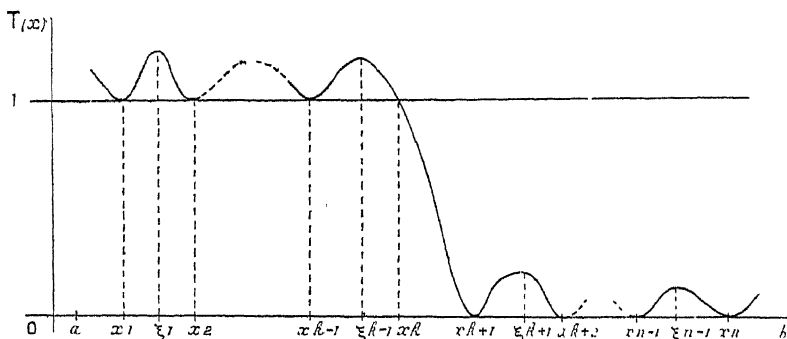



Fig. 27

Il reste à montrer que l'on peut calculer effectivement ce polynome $T(x)$, à partir des conditions (13) et (14). Il suffit d'appliquer la formule d'interpolation de Lagrange, en considérant la connaissance de $T(x_1)$ et de $T'(x_1)$, par exemple, comme le cas limite de la connaissance de $T(x_1)$ et de $T(x'_1)$, x'_1 étant infiniment voisin de x_1.

Pour voir ce que devient la formule d'interpolation dans cette circonstance, désignons par $x'_1, x'_2, \ldots, x'_{k-1}, x'_{k+1}, \ldots, x'_n$ des abscisses infiniment voisines de $x_1, x_2, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n$, et posons

$$\varpi(x) = (x-x_1)(x-x'_1)\ldots(x-x_{k-1})(x-x'_{k-1})$$
$$(x-x_k)(x-x_{k+1})(x-x'_{k+1})\ldots(x-x_n)(x-x'_n).$$

Si l'on se donnait les valeurs de $T(x)$ pour tous les zéros de $\varpi(x)$, la formule de Lagrange s'écrirait

$$\frac{T(x)}{\varpi(x)} = \frac{T(x_1)}{(x-x_1)\varpi'(x_1)} + \frac{T(x'_1)}{(x-x'_1)\varpi'(x'_1)} + \ldots + \frac{T(x'_n)}{(x-x'_n)\varpi'(x'_n)}.$$

Si nous posons

$$\varpi(x) = (x-x_1)(x-x'_1)\varpi_1(x)$$

nous avons

$$\varpi'(x_1) = (x_1 - x'_1)\varpi_1(x_1),$$
$$\varpi'(x'_1) = (x'_1 - x_1)\varpi_1(x'_1),$$

et il vient

$$\frac{T(x)}{\varpi(x)} = \frac{1}{(x'_1 - x_1)} \left\{ \frac{-T(x_1)}{(x-x_1)\varpi_1(x_1)} + \frac{T(x'_1)}{(x-x'_1)\varpi_1(x'_1)} \right\} + \ldots$$

Si maintenant nous faisons tendre x'_1 vers x_1, le terme écrit dans le second membre a une limite bien déterminée

$$\frac{d}{dx_1}\left[\frac{T(x_1)}{(x-x_1)\varpi_1(x_1)}\right] = \frac{T'(x_1)\varpi_1(x_1) - T(x_1)\varpi'_1(x_1)}{(x-x_1)\varpi_1(x_1)^2} + \frac{T(x_1)}{(x-x_1)^2\varpi_1(x_1)}.$$

D'une manière générale, lorsque tous les x'_i tendent vers les x_i $(i = 1, 2, \ldots, k-1, k+1, \ldots, n)$, $\varpi(x)$ devient

$$S(x) = (x-x_1)^2 \ldots (x-x_{k-1})^2 (x-x_k)(x-x_{k+1})^2 \ldots (x-x_n)^2,$$

et nous obtenons le développement

$$\frac{T(x)}{S(x)} = \frac{B_1}{(x-x_1)^2} + \frac{C_1}{x-x_1} + \ldots$$
$$+ \frac{B_{k-1}}{(x-x_{k-1})^2} + \frac{C_{k-1}}{x-x_{k-1}} + \frac{C_k}{x-x_k} + \frac{B_{k+1}}{(x-x_{k+1})^2} + \ldots + \frac{C_n}{x-x_n},$$

dont nous savons calculer tous les coefficients B_i, C_i, en fonction des quantités $T(x_i)$ et $T'(x_i)$.

21. Le polynome $T(x)$ ayant été ainsi déterminé, et vérifiant les conditions que nous lui avons imposées, appliquons-lui l'identité (7). Elle s'écrit

$$I = \int_a^b T(x)\,df(x) = a_1 + a_2 + \ldots + a_k.$$

Or, cette intégrale peut être décomposée en deux :

$$I = \int_a^{x_k} T(x)\,df(x) + \int_{x_k}^b T(x)\,df(x),$$

et chacune de ces deux intégrales est positive ; d'autre part, $T(x)$ est constamment supérieur ou égal à l'unité dans la première intégrale, de sorte que l'on peut écrire

$$\int_a^{x_k} df(x) < \int_a^{x_k} T(x)\,df(x) < I,$$

et, par suite,

$$\int_a^{x_k} df(x) < a_1 + a_2 + \ldots + a_k, \tag{15}$$

et cette inégalité est satisfaite quelle que soit la racine x_k. On aurait pu, dans le raisonnement, permuter les rôles de a et de b, ce qui aurait conduit à l'inégalité analogue

$$\int_{x_{k+1}}^b df(x) < a_n + a_{n-1} + \ldots + a_{k+1},$$

qui, retranchée de l'égalité

$$\int_a^b df(x) = a_1 + a_2 + \ldots + a_n,$$

entraîne

$$\int_a^{x_{k+1}} df(x) > a_1 + a_2 + \ldots + a_k. \tag{16}$$

Nous obtenons ainsi, en (15) et en (16), une double inégalité très remarquable :

$$\int_a^{x_k} df(x) < a_1 + a_2 + \ldots + a_k < \int_a^{x_{k+1}} df(x); \tag{17}$$

c'est elle que nous avions en vue en choisissant le polynome particulier $T(x)$, et elle va nous permettre des conclusions de la plus haute importance.

22. Traçons sur une même figure la courbe $y = f(x) = \int_a^x df(x)$ et la courbe en escalier relative à $Q_n(x)$. Nous allons voir que les

n marches de cette dernière sont traversées, comme l'indique la figure, par la première courbe $y=f(x)$

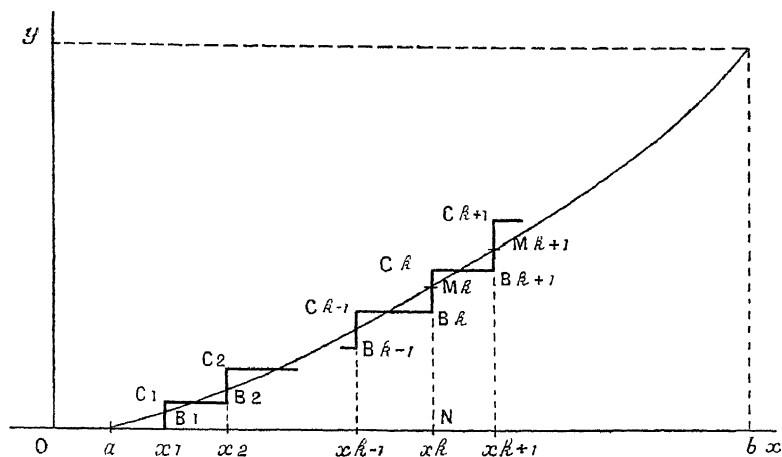



Fig. 28.

En effet, si $f(x_k)$ est représentée par l'ordonnée NM_k, le point M_k, d'après (17), est compris entre les deux extrémités B_k et C_k de la marche d'abscisse x_k; d'une manière plus précise, la double inégalité

$$NB_k < NM_k < NC_k$$

se traduit par

$$a_1 + a_2 + \ldots + a_{k-1} < f(x_k) < a_1 + a_2 + \ldots + a_k$$

Il est clair, d'ailleurs, que ceci ne suppose pas necessairement que $y=f(x)$ est representable par une courbe ordinaire.

On voit ainsi que cet escalier est une véritable approximation de la fonction $y=f(x)$.

Supposons alors que n augmente indéfiniment, nous allons voir que les hauteurs des marches de l'escalier tendent toutes vers zéro, de sorte que l'approximation qu'il nous fournit de $f(x)$ sera de plus en plus satisfaisante; dans ces conditions, la fonction proposée $y=f(x)$ apparaît comme la limite de l'escalier défini par $Q_n(x)$, lorsque n augmente indéfiniment.

Pour vérifier notre proposition, nous supposerons que l'intervalle (a, b) est fini, et nous allons montrer que, dans tout intervalle (α, β) de (a, b), on peut trouver une racine au moins de tout

polynome $Q_n(x)$, dès que le rang de ce polynome est suffisamment grand.

Soit, en effet, (α', β') un intervalle intérieur à (α, β). Cherchons une fonction $\varphi(x)$, définie et continue entre a et b, qui vérifie les inégalités suivantes :

$$\begin{cases} 0 < \varphi(x) < \varepsilon & \text{pour } a < x < \alpha \text{ et } \beta < x < b, \\ 0 < \varphi(x) \leq 1 & \text{pour } \alpha < x < \alpha' \text{ et } \beta' < x < \beta, \\ 1 \leq \varphi(x) & \text{pour } \alpha' \leq x \leq \beta'. \end{cases}$$

Il est évident que, ε étant donné, on peut toujours trouver une telle fonction $\varphi(x)$. D'autre part, d'après le théorème bien connu de Weierstrass sur l'approximation des fonctions, on sait trouver un polynome $T(x)$, tel que, dans l'intervalle (a, b), on ait

$$|\varphi(x) - T(x)| < \varepsilon.$$

$T(x)$ satisfait alors à des inégalités de même nature que $\varphi(x)$, la seule modification étant le changement possible de ε en 2ε. Désignons

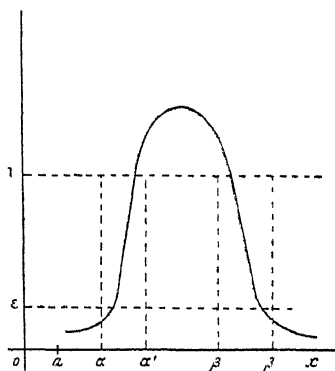



Fig. 29.

par k le degré de ce polynome. Donnons à n une valeur assez grande pour que $2n - 2$ soit supérieur à k, et considérons le polynome $Q_n(x)$, si $x_1, x_2, \ldots, x_n$ sont ses zéros, $a_1, a_2, \ldots, a_n$ les masses qui y sont réparties, la formule générale (7) est applicable et nous fournit la valeur de l'intégrale

$$J = \int_a^b T(x)\, df(x) = a_1 T(x_1) + a_2 T(x_2) + \ldots + a_n T(x_n).$$

Si aucun des zéros de $Q_n(x)$ ne se trouvait dans l'intervalle (α, β), tous les $T(x_i)$ seraient inférieurs à 2ε, et l'on aurait

$$J < 2\varepsilon(a_1 + a_2 + \ldots + a_n),$$

ou

$$J < 2\varepsilon c_0.$$

Mais, d'autre part, si $df(x)$ n'est pas constamment nul entre α et β, nous pouvons choisir l'intervalle (α', β') de façon que l'intégrale $\int_{\alpha'}^{\beta'} df(x)$ ne soit pas nulle, de sorte que nous aurions

$$0 < \int_{\alpha'}^{\beta} df(x) < \int_a^b T(x)\,df(x) < 2\varepsilon c_0.$$

Cette triple inégalité est contradictoire lorsque n augmente indéfiniment, car le deuxième membre est fini et non nul, alors que le dernier est infiniment petit

Si, entre α et β, la fonction $f(x)$ était constante, notre démonstration n'aurait plus de valeur. Mais on démontre alors qu'il y a une racine x_i dans tout intervalle qui contient le palier de $f(x)$

D'une façon générale, on peut dire que, dans tout intervalle (α, β) où $f(x)$ a une variation non nulle, on peut placer une marche d'un polynome $Q_n(x)$.

Enfin, si nous nous reportons à la double inégalité

$$\int_a^{x_k} df(x) < a_1 + a_2 + \ldots + a_k < \int_a^{x_{k+1}} df(x),$$

nous avons également

$$\int_a^{x_{k-1}} df(x) < a_1 + a_2 + \ldots + a_{k-1} < \int_a^{x_k} df(x),$$

ce qui nous donne, en retranchant les membres médians,

$$a_k < \int_{x_{k-1}}^{x_{k+1}} df(x)$$

Il résulte immédiatement de la combinaison de ces deux résultat que a_k tend vers zéro, avec l'intervalle (x_{k-1}, x_{k+1}) lorsque n augmente indéfiniment.

Résumons, en quelques mots, les résultats les plus importants de

cette étude : « Étant donnée une fonction arbitraire $f(x)$, croissante dans un intervalle fini et positif (a, b), les $2n$ premiers moments de cette fonction appartiennent à une fonction en escalier, de n marches. Cet escalier fournit une approximation de la fonction initiale $f(x)$, et a pour limite cette fonction lorsque le nombre de ses marches augmente infiniment. »

Telle est la conclusion à laquelle nous voulions arriver, et c'est par elle que nous terminerons cet exposé.

NOTE I.

SUR LES VALEURS MOYENNES

1 La note ci-dessous est le résumé d'un article de Tchebychef, traduit dans le *Journal de Liouville* de l'année 1867. Dans cet article, l'auteur considère un certain nombre de variables $x, y, z, \ldots$, dont il suppose connues les valeurs moyennes, ainsi que les valeurs moyennes de leurs carrés Soient

$$\overline{x} = a, \quad \overline{y} = b, \quad \overline{z} = c, \ldots$$
$$\overline{x^2} = a_1 \quad \overline{y^2} = b_1, \quad \overline{z^2} = c_1, \ldots$$

ces valeurs moyennes.

Il se propose de déterminer une limite inférieure de la probabilité pour que l'écart u de la somme de ces quantités variables, c'est-à-dire la différence entre cette somme et sa valeur moyenne, soit compris entre certaines limites

Auparavant formons la moyenne quadratique de cet écart, elle est égale

$$\overline{u^2} = \overline{(x + y + z + \quad - a - b - c \ldots)^2},$$

ou encore

$$\overline{u^2} = \overline{(x - a)^2} + \overline{(y - b)^2} + \overline{(z - c)^2} + \ldots$$

car les moyennes des écarts $(x - a)$, $(y - b)$, $(z - c)$, ..., sont nulles.

Le développement du second membre donne

$$\overline{u^2} = \overline{x^2} - 2\overline{ax} + a^2 + \ldots,$$

ou encore

$$\overline{u^2} = a_1 + b_1 + c_1 + \ldots - (a^2 + b^2 + c^2 + \ldots$$

Designons par P la probabilité pour que u^2 ne dépasse pas une quantité de la forme $\alpha^2 \overline{u^2}$, α étant un nombre que l'on suppose plus grand que 1. La probabilité complémentaire $P' = 1 - P$ est la probabilité pour que u^2 soit superieur à $\alpha^2 \overline{u^2}$; il est facile de voir que P' est nécessairement inférieur à $\frac{1}{\alpha^2}$.

Pour cela, reprenons la définition de la valeur moyenne $\overline{u^2}$; c'est la somme

$$\sum_i p_i u_i^2,$$

p_i étant la probabilité pour que u^2 prenne la valeur u_i^2, et cette somme étant étendue à toutes les valeurs possibles de u^2. Cette valeur moyenne surpasse donc la somme $\Sigma p_i u_i^2$, étendue aux seules valeurs $u_i^2 > \alpha^2 \overline{u^2}$, elle est donc plus grande, *a fortiori*, que

$$\alpha^2 \overline{u^2} \Sigma p_i,$$

Σp_i désignant alors la probabilité pour que u^2 soit supérieur à $\alpha^2 \overline{u^2}$.

En définitive, on peut écrire

$$\overline{u^2} > \alpha^2 \overline{u^2} \mathrm{P}',$$

qui entraîne bien l'inégalité énoncée $\mathrm{P}' < \frac{1}{\alpha^2}$; remarquons d'ailleurs que cette inégalité n'offre quelque intérêt que si α est supérieur à 1, comme nous l'avons supposé.

Il résulte de là que P est supérieur à $1 - \frac{1}{\alpha^2}$, ce qui s'exprime par le théorème suivant

Théorème. — *La probabilité pour que la somme des quantités variables $x, y, z, \ldots$ soit comprise entre*

$$a + b + c + \ldots - \alpha\sqrt{a_1 + b_1 + c_1 + \ldots - (a^2 + b^2 + c^2 + \ldots)}$$

et

$$a + b + c + \ldots + \alpha\sqrt{a_1 + b_1 + c_1 + \ldots - (a^2 + b^2 + c^2 + \ldots)}$$

est toujours plus grande que $1 - \frac{1}{\alpha^2}$, quel que soit α

2. Supposons que ces variables soient au nombre de n. On peut introduire leur moyenne arithmétique ainsi que les moyennes arithmétiques de leurs valeurs moyennes. La condition de limitation de l'écart peut s'écrire, en effet,

$$\left| \frac{x + y + z + \ldots}{n} - \frac{a + b + c + \ldots}{n} \right| < \frac{\alpha}{\sqrt{n}} \sqrt{\frac{a_1 + b_1 + c_1 + \ldots}{n} - \frac{a^2 + b^2 + c^2 + \ldots}{n}};$$

posons $\alpha = \frac{\sqrt{n}}{t}$; le théorème précédent exprime alors que la probabilité pour que

$$\left| \frac{x + y + z + \ldots}{n} - \frac{a + b + c + \ldots}{n} \right|$$

ne surpasse pas

$$\frac{1}{t} \sqrt{\frac{a_1 + b_1 + \ldots}{n} - \frac{a^2 + b^2 + \ldots}{n}}$$

est toujours plus grande que $1 - \frac{t^2}{n}$, quel que soit t.

On peut déduire de ce résultat une conclusion intéressante dans le cas où la moyenne $\frac{a_1+b_1+c_1+\ldots}{n}$ est finie, quelque grand que soit le nombre n des quantités $x, y, z, \ldots$ choisies dans une suite infinie. La quantité sous le radical est, *a fortiori*, bornée supérieurement, et si M est sa borne supérieure, il résulte du théorème que la probabilité pour que la différence

$$\frac{x+y+z+\ldots}{n}-\frac{a+b+c+\ldots}{n}$$

soit inférieure à $\frac{\sqrt{M}}{t}$, en valeur absolue, est toujours plus grande que $1-\frac{t^2}{n}$.

Or, on peut rendre $\frac{\sqrt{M}}{t}$ aussi petit que l'on veut, en donnant à t une valeur suffisamment grande, on peut simultanément faire croître n indéfiniment de façon que $\frac{t^2}{n}$ tende vers zéro, dans ces conditions P tend vers 1, ce qui conduit au théorème suivant :

THÉORÈME. — *Si la moyenne arithmétique des valeurs moyennes quadratiques de quantités variables, prises en nombres de plus en plus grands, reste bornée supérieurement, la probabilité que la différence, entre la moyenne arithmétique de ces quantités et celle de leurs valeurs moyennes, soit moindre en valeur absolue qu'une quantité donnée, aussi petite que l'on veut, tend vers l'unité lorsque le nombre des quantités considérées augmente indéfiniment.*

3. Ce théorème conduit immédiatement à une généralisation du théorème de Bernoulli. Si, en effet, $P_1, P_2, P_3, \ldots$ sont les probabilités d'un événement dans une succession d'épreuves $E_1, E_2, E_3, \ldots$, on peut donner à x la valeur 1 si cet événement se produit dans l'épreuve E_1, sinon la valeur 0 et faire la même convention pour $y, z, \ldots$ relativement aux épreuves $E_2, E_3, \ldots$.

Dans ces conditions, $\frac{x+y+z+\ldots}{n}$ représente la fréquence de l'événement; d'autre part $\frac{a+b+c+\ldots}{n}$ et $\frac{a_1+b_1+c_1+\ldots}{n}$ sont égaux à $\frac{P_1+P_2+P_3+\ldots}{n}$, donc

Lorsque le nombre d'épreuves devient infini, on obtient une probabilité, infiniment voisine de 1, que la différence entre la fréquence de cet événement et sa probabilité moyenne soit moindre que toute quantité donnée.

On retrouve bien le théorème de Bernoulli si la probabilité de l'événement est la même dans toutes les épreuves.

4. **Notes bibliographiques.** — Ce théorème fondamental de Tchebychef

a donné lieu à de très nombreux travaux et à diverses généralisations. Citons, en particulier

MARKOFF *Calcul des Probabilités* (édition russe, Saint-Pétersbourg 1913)

Karl PEARSON *Biometrika*

CANTELLI *Rendiconti della R Accademia dei Lincei*, 1916

On pourra consulter également à ce sujet les Notes récentes suivantes parues aux *Comptes rendus de l'Académie des Sciences*, durant l'année 1922

Alf GULDBERG α *Sur le théorème de Tchebychef* (4 septembre 1922) Dans cette Note, l'auteur introduit les valeurs moyennes des puissances autres que le carré

β *Sur un théorème de M. Markoff* (23 octobre 1922)

γ *Sur les valeurs moyennes* (27 novembre 1922). Cette Note étend les théorèmes de Tchebychef et Markoff aux valeurs moyennes relatives

δ *Sur quelques inégalités du calcul des probabilités* (26 décembre 1922)

Constant LURQUIN . *Sur le critérium de Tchebychef* (24 octobre 1922).

Birger MEIDELL *Sur un problème du calcul des probabilités et les statistiques mathématiques* (6 novembre 1922)

NOTE II.

SUR LES POLYNOMES D'HERMITE-TCHEBYCHEF.

1. Les questions étudiées par Stieltjes, et dont nous avons donné un aperçu dans le dernier Chapitre, ont également été étudiées par le mathématicien russe Tchebychef. Ses démonstrations sont plus longues que celles de Stieltjes, et même parfois pénibles, mais elles s'appliquent au problème des moments sous sa forme la plus générale, et, d'ailleurs, elles ont été grandement améliorées depuis.

Parmi les questions particulièrement intéressantes que traitent les travaux de ce géomètre, nous allons dire quelques mots des polynomes $Q_n(x)$ relatifs à la fonction de Gauss

$$f(x) = \int_{-\infty}^{x} e^{-x^2}\,dx$$

L'intervalle (a, b) est ici l'intervalle de variation de x, $(-\infty, +\infty)$. Ces polynomes sont les polynomes d'Hermite-Tchebychef.

On est encore conduit à les considérer dans l'expression des dérivées successives de la fonction $y = \dfrac{df(x)}{dx} = e^{-x^2}$, ces dérivées sont, en effet,

$$\begin{aligned} y' &= -2x\,e^{-x^2}, \\ y'' &= (4x^2 - 2)\,e^{-x^2}, \\ y''' &= (-8x^3 + 12x)\,e^{-x^2}, \\ \ldots & \quad \ldots \end{aligned}$$

et, d'une manière générale, on peut poser

$$y^{(n)} = (-2)^n\, Q_n(x)\, e^{-x^2}.$$

Le polynome $Q_n(x)$ est de degré n, et le coefficient de son terme de plus haut degré est égal à l'unité ; on voit, de plus, que tous les termes sont de même parité.

Ce sont justement les polynomes d'Hermite-Tchebychef. Pour le voir, il suffit de montrer que l'intégrale

$$\int_{-\infty}^{+\infty} Q_n(x)\, F(x)\, df(x)$$

est nulle pour tout polynome $F(x)$ de degré $n - 1$ au plus.

Posons

$$I = \int_{-\infty}^{+\infty} Q_n(x)\, G(x)\, e^{-x^2}\, dx$$

Cette intégrale, d'après la définition de $Q_n(x)$, s'écrit encore

$$I = \frac{(-1)^n}{2^n} \int_{-\infty}^{+\infty} \frac{d^n e^{-x^2}}{dx^n} G(x)\, dx$$

Une première intégration par parties donne

$$I = \frac{(-1)^{n-1}}{2^n} \int_{-\infty}^{+\infty} \frac{d^{n-1} e^{-x^2}}{dx^{n-1}} G'(x)\, dx,$$

qui est de la même forme que l'intégrale primitive. On peut effectuer successivement n intégrations par parties, ce qui fournit pour I l'expression définitive

$$I = \frac{1}{2^n} \int_{-\infty}^{+\infty} e^{-x^2}\, G^{(n)}(x)\, df(x)$$

Cette expression montre immédiatement que I est nul si $G(x)$ est un polynome $F(x)$ de degré $n-1$ au plus.

2. Ce que nous avons dit, en général, dans le Chapitre VI, des polynomes $Q_n(x)$ peut être répété ici avec certaines simplifications. La formule (5) de ce Chapitre

$$Q_n(x) = (x - \alpha_{n-1})\, Q_{n-1}(x) - \lambda_{n-1} Q_{n-2}(x),$$

est encore valable; mais ici

$$\alpha_{n-1} = \frac{\displaystyle\int_{-\infty}^{+\infty} x\, Q_{n-1}(x)^2 e^{-x^2}\, dx}{\displaystyle\int_{-\infty}^{+\infty} Q_{n-1}(x)^2 e^{-x^2}\, dx}$$

est nul, et

$$\lambda_{n-1} = \frac{\displaystyle\int_{-\infty}^{+\infty} Q_{n-1}(x)^2 e^{-x^2}\, dx}{\displaystyle\int_{-\infty}^{+\infty} Q_{n-2}(x)^2 e^{-x^2}\, dx}.$$

Pour calculer λ_{n-1}, calculons l'intégrale générale

$$\int_{-\infty}^{+\infty} Q_n(x)^2 e^{-x^2}\, dx.$$

C'est l'intégrale I, dans laquelle on choisit pour $G(x)$ le polynome $Q_n(x)$ lui-même, ce qui donne

$$\int_{-\infty}^{+\infty} Q_n(x)^2 e^{-x^2} dx = \frac{1}{2^n}\int_{-\infty}^{+\infty} e^{-x^2} \frac{d^n Q_n(x)}{dx^n} dx = \frac{n!}{2^n}\int_{-\infty}^{+\infty} e^{-x^2} dx = \frac{n!}{\sqrt{\pi}.2^n},$$

et, par suite,

$$\lambda_{n-1} = \frac{n-1}{2}.$$

Les polynomes $Q_n(x)$ satisfont donc à la relation de récurrence

$$Q_n(x) = x\,Q_{n-1}(x) - \frac{n-1}{2}\,Q_{n-2}(x).$$

On peut également arriver à cette formule en partant de la définition de ces polynomes

$$Q_n(x) = \frac{(-1)^n}{2^n}\, e^{x^2}\, \frac{d^n e^{-x^2}}{dx^n}.$$

Nous pouvons écrire, en effet,

$$\begin{aligned}\frac{d^n e^{-x^2}}{dx^n} &= \frac{d}{dx}\left[(-2)^{n-1} Q_{n-1}(x) e^{-x^2}\right]\\ &= (-2)^n x Q_{n-1}(x) e^{-x^2} + (-2)^{n-1} Q'_{n-1}(x) e^{-x^2}\end{aligned}$$

d'où l'on tire

$$Q_n(x) = x\,Q_{n-1}(x) - \frac{1}{2}\,Q'_{n-1}(x).$$

Pour vérifier que $Q'_{n-1}(x)$ est égal à $(n-1)\,Q_{n-2}(x)$, il suffit de multiplier les deux membres de cette dernière relation par $F(x)e^{-x^2}dx$, $F(x)$ étant un polynome arbitraire de degré $n-3$ au plus, et d'intégrer entre $-\infty$ et $+\infty$; on a bien

$$\int_{-\infty}^{+\infty} Q'_{n-1}(x)\,F(x)\,e^{-x^2}dx = 0,$$

ce qui caractérise les polynomes proportionnels à $Q_{n-2}(x)$, la valeur de la constante $\dfrac{Q'_{n-1}(x)}{Q_{n-2}(x)}$ est enfin égale au rapport des termes de degré $n-2$ de ces deux polynomes, c'est-à-dire à $(n-1)$.

3. Au polynome $Q_n(x)$ correspond une fonction en escalier qui admet les mêmes $2n$ premiers moments que la fonction de Gauss. Les masses $a_1, a_2, \ldots, a_n$ réparties aux zéros $x_1, x_2, \ldots, x_n$ de $Q_n(x)$ sont données par la formule générale

$$a_i = \frac{A}{Q_{n-1}(x_i)\,Q'_n(x_i)},$$

où A est l'intégrale $\int_{-\infty}^{+\infty} Q_{n-1}(x)^2\, e^{-x^2} dx$ dont nous connaissons la

valeur $\frac{(n-1)!}{\sqrt{\pi}.2^{n-1}}$. Enfin, en remplaçant $Q'_n(x_i)$ par $nQ_{n-1}(x_i)$, nous obtenons pour ces masses, l'expression

$$\frac{(n-1)!}{\sqrt{\pi}\,2^{n-1}\,n\,Q_{n-1}(x_i)^2}.$$

Cette formule va nous permettre de montrer que les a_i tendent tous vers zéro lorsque n augmente indéfiniment. Pour cela nous allons déterminer une limite inférieure de la valeur de $Q_{n-1}(x_i)^2$.

Considérons la fonction

$$z = Q_n(x)^2 + \lambda Q_{n-1}(x)^2,$$

où λ est une constante dont nous choisirons la valeur; la dérivée de cette fonction peut s'écrire

$$\frac{dz}{dx} = Q_{n-1}(x)\left[2nxQ_{n-1}(x) + (2\lambda - n)(n-1)Q_{n-2}(x)\right],$$

et se simplifie beaucoup si l'on prend λ égal à $\frac{n}{2}$.

Il vient alors

$$\frac{dz}{dx} = 2nxQ_{n-1}(x)^2,$$

de sorte que cette fonction z particulière a un minimum absolu pour $x = 0$; c'est ce minimum

$$Q_n(0)^2 + \frac{n}{2}Q_{n-1}(0)^2$$

que l'on prend pour limite inférieure de la valeur de $z(x_i) = \frac{n}{2}Q_{n-1}(x_i)^2$.

Pour calculer cette limite inférieure, remarquons que la relation de récurrence donne, pour $x = 0$,

$$Q_n(0) = -\frac{n-1}{2}Q_{n-2}(0),$$

et, en particulier,

$$|Q_{2p}(0)| = \frac{1.3\ldots(2p-1)}{2^p}Q_0(0) = \frac{1.3\ldots(2p-1)}{2^p}.$$

D'autre part, $Q_{2p+1}(0)$ est nul, d'où résulte, enfin, pour $n = 2p+1$,

$$\frac{2p+1}{2}Q_{2p}(x_i)^2 > \frac{2p+1}{2}Q_{2p}(0)^2 = (2p+1)\frac{[1.3\ldots(2p-1)]^2}{2^{2p+1}}.$$

Supposons que n soit un nombre impair $2p+1$, et remplaçons, dans l'expression de a_i, $Q_{n-1}(x_i)^2$ par cette limite inférieure; nous obtenons, pour a_i, une limite supérieure

$$\frac{(2p)!}{\sqrt{\pi}(2p+1)\,[1.3\ldots(2p-1)]^2},$$

ou enfin

$$a_l < \frac{1}{\sqrt{\pi}} \cdot \frac{2.4 \ldots 2p}{3.5 \ldots 2p+1};$$

le second membre de cette inégalité étant infiniment petit lorsque p augmente indéfiniment, il en est de même de a_l. Un raisonnement analogue conduirait à la même conclusion, si n était pair.

Enfin, les courbes en escalier successives sont des approximations de plus en plus grandes de la courbe $y = \int_{-\infty}^{x} e^{-x^2} dx$, et dans tout intervalle fini, convergent uniformement vers cette courbe. Nous ne donnerons pas la démonstration de ce résultat (1), qui a été généralisé et simplifié dans un Mémoire récent de M. Polya, dont nous disons quelques mots un peu plus loin.

---

(1) Sur ce résultat, cf. J. MARKOFF, *Calcul des Probabilités* (Saint-Pétersbourg, 1913) « Démonstration du deuxième théoreme limite du calcul des probabilités par la méthode des moments ».

NOTES BIBLIOGRAPHIQUES

SUR LE PROBLÈME DES MOMENTS

1. Parmi les travaux de Tchebychef, citons encore les recherches suivantes.

On a vu au Chapitre VI de cet ouvrage que si l'on désigne par $c_0, c_1, c_2, \ldots$ les moments d'une répartition effective $df(x)$, les déterminants

$$\Delta_k = \begin{vmatrix} c_0 & c_1 & \ldots & c_k \\ c_1 & c_2 & \ldots & c_{k+1} \\ & & & \ldots \\ c_k & c_{k+1} & \ldots & c_{2k} \end{vmatrix}$$

sont tous positifs, et l'on a démontré, en même temps, que les racines de l'équation

$$Q_n(x) = \frac{1}{\Delta_{n-1}} \begin{vmatrix} 1 & x & x^2 & \ldots & x^n \\ c_0 & c_1 & c_2 & \ldots & c_n \\ \cdot & \cdot\cdot & \cdot & \cdot\cdot & \cdot\cdot \\ c_{n-1} & c_n & c_{n+1} & \ldots & c_{2n-1} \end{vmatrix} = 0$$

sont toutes réelles.

Tchebychef démontre que la seule condition, que tous les Δ_k sont positifs, suffit pour que toutes ces racines soient réelles. On peut mettre $Q_n(x)$ sous la forme

$$Q_n(x) = \frac{1}{\Delta_{n-1}} \begin{vmatrix} c_1 - c_0 x & c_2 - c_1 x & \ldots & c_n - c_{n-1} x \\ c_2 - c_1 x & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ c_n - c_{n-1} x & \ldots & \ldots & c_{2n-1} - c_{2n-2} x \end{vmatrix};$$

ce déterminant apparaît ainsi comme le discriminant de la forme quadratique

$$\varphi(z_1, z_2, \ldots, z_n) - x\psi(z_1, z_2, \ldots, z_n),$$

où l'on pose

$$\varphi(z_1, z_2, \ldots, z_n) = \sum_{i=1}^{n} \sum_{k=1}^{n} c_{i+k-1} z_i z_k$$

et

$$\psi(z_1, z_2, \ldots, z_n) = \sum_{i=1}^{n} \sum_{k=1}^{n} c_{i-k-2} z_i z_k$$

L'auteur démontre alors que ψ est une forme définie positive. L'équation en S des deux formes φ et ψ a bien alors toutes ses racines réelles, d'après la théorie classique des formes quadratiques.

Si la forme φ était elle-même définie positive, les zéros de $Q_n(x)$ seraient, de plus, tous positifs.

2. Le problème des moments, de Stieltjes et Tchebychef, a donné lieu récemment à un Mémoire très important de M. Polya [1], intitulé : *Ueber den zentralen Greuzwertsatz der Wahrscheinlichkeitsrechnung und das Moment-Problem*.

L'auteur considère des fonctions monotones, continues à droite, sans l'être nécessairement à gauche, définies de $-\infty$ à $+\infty$, et vérifiant les deux conditions

$$f(-\infty) = 0, \qquad f(+\infty) = 1.$$

Ce sont les « fonctions de répartition », telles que nous les avons définies dans la première partie du Chapitre VI.

Il démontre tout d'abord le

THÉORÈME I. — *Si une suite de fonctions de répartition $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ converge vers une fonction de répartition continue, $f(x)$, elle converge uniformément vers cette fonction limite.*

Pour démontrer ce théorème, il suffit de partager l'intervalle de variation $(-\infty, +\infty)$ en un nombre fini d'intervalles partiels,

$$-\infty \quad A_1 \quad A_2 \quad \ldots \quad A_p \quad +\infty,$$

tels que, dans chacun d'eux, la croissance de $f(x)$ soit inférieure à un nombre donné d'avance ε.

On peut évidemment trouver un nombre N tel que $n > N$ entraîne

$$|f_n(A_i) - f(A_i)| < \varepsilon,$$

en tous les points A_i, puisque ces points sont en nombre fini, cette inégalité est vraie *a fortiori* aux limites extrêmes $-\infty$ et $+\infty$, que l'on peut désigner par A_0 et A_{p+1}.

Dans ces conditions, une valeur quelconque de x est comprise dans un intervalle (A_{i-1}, A_i) et la condition de croissance des fonctions $f_n(x)$ et $f(x)$

(1) Cf. *Mathematische Zeitschrift*, 1920.

donne lieu à la double inégalité

$$f_n(x) - f(x) \leqq f_n(A_i) - f(A_{i-1}) = f_n(A_i) - f(A_i) + f(A_i) - f(A_{i-1}) \leqq 2\varepsilon,$$
$$f_n(x) - f(x) \geqq f_n(A_{i-1}) - f(A_i) = f_n(A_{i-1}) - f(A_{i-1}) + f(A_{i-1}) - f(A_i) \geqq -2\varepsilon.$$

Cette double inégalité, qui s'écrit encore

$$|f_n(x) - f(x)| \leqq 2\varepsilon,$$

exprime bien que la convergence est uniforme dans tout l'intervalle $(-\infty, +\infty)$.

Citons encore, sans les démontrer, les deux théorèmes suivants, du même Mémoire.

Théorème II. — *Étant données des fonctions continues* $F_n(x)$ *dont les dérivées à droite sont des fonctions de répartition, si la suite* $F_1(x)$, $F_2(x)$, ..., $F_n(x)$, ... *converge vers une fonction dérivable* $F(x)$, *dont la dérivée est également une fonction de répartition, la suite des dérivées à droite des fonctions* $F_n(x)$ *converge uniformément vers* $F'(x)$.

Théorème III. — *Si une suite de fonctions de répartition* $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, ... *est telle que les moments de* $f_n(x)$ *tendent vers les moments*

$$\mu_m = \int_{-\infty}^{+\infty} x^m \, df(x)$$

d'une fonction de répartition $f(x)$, *continue, cette suite converge vers la fonction* $f(x)$, *et la convergence est uniforme dans tout intervalle.*

Remarquons que ce dernier théorème généralise le théorème de Tchebychef et Markoff relatif à la fonction de Gauss

$$f(x) = \frac{1}{\pi} \int_{-\infty}^{x} e^{-x^2} \, dx,$$

et que nous n'avions fait qu'énoncer à la fin de la Note 2.

3. Enfin, sur le même problème, on pourra encore consulter les deux Mémoires suivants :

Hamburger, *Ueber eine Erweiterung des Stieltjesschen Momentenproblems* (*Mathematische Annalen*, 1920-1921), et, plus récemment, Torsten Carleman, *Sur le problème des moments* (*C. R. Ac. Sciences*, 26 juin 1922).

TABLE DES MATIÈRES

Pages

CHAPITRE I. — *Généralités* 1

CHAPITRE II. — *Problèmes du premier ordre* 6

Probabilités discontinues 6
Probabilités continues 17
Probabilités dénombrables 24

CHAPITRE III. — *Probabilités discontinues. Problèmes du deuxième ordre* 29

Loi de Gauss 29
Problème du deuxième ordre 41

CHAPITRE IV. — *Probabilités continues. Problèmes du deuxième ordre* 57

Problèmes sur la droite 57
Problèmes dans l'espace 74

CHAPITRE V. — *Jeu de pile ou face* 91

CHAPITRE VI. — *Statistique* 105

Généralités sur les fonctions de la statistique 105
Problème des moments 119

NOTES

NOTE I. — *Sur les valeurs moyennes* 147

NOTE II. — *Sur les polynomes d'Hermite-Tchebychef* 151

Notes bibliographiques sur le problème des moments 156

TRAITÉ DU CALCUL DES PROBABILITÉS ET DE SES APPLICATIONS

Par Émile BOREL

Professeur de Calcul des Probabilités et de Physique mathématique à la Faculté des Sciences de Paris,
Membre de l'Institut

TOME I — *Les principes de la Théorie des probabilités*

1 Principes et formules classiques, par Émile BOREL, rédigé par R LAGRANGE, 160 pages, 29 figures . 25 fr
2 Erreurs et moindres carrés, par Robert DELTHEIL
3 Recherches théoriques modernes, par Maurice FRÉCHET
4. Les principes de la statistique mathématique, par CH RISSER et C -E. TRAYNARD

TOME II. — *Les applications de la Théorie des probabilités aux sciences mathématiques et aux sciences physiques*

1 Applications à l'arithmétique et à la théorie des fonctions, par Émile BOREL, rédigé par Paul DUBREIL.. ..... 25 fr
2 Probabilités géométriques, par Robert DELTHEIL 30 fr
3 Mécanique statistique classique, par Émile BOREL, rédigé par Francis PERRIN, 148 pages, 13 figures . . .. 25 fr
4 Applications à l'astronomie, par C.-V -L CHARLIER
5 Applications aux théories physiques actuelles, par Francis PERRIN

TOME III — *Les applications de la Théorie des probabilités aux sciences économiques et aux sciences biologiques*

1. Assurances sur la vie Calcul des primes, par Henri GALBRUN, 210 pages. . . . . 45 fr.
2. Assurances sur la vie Calcul des réserves, par Henri GALBRUN 40 fr
3 Applications à la biologie. Variations continues et Biométrie, par CH. RISSER et C -E. TRAYNARD.

TOME IV. — *Applications diverses et conclusion*

1. Applications au tir, par J HAAG . . .. 35 fr.
2. Applications aux jeux de hasard, par Émile BOREL.
3 Compléments divers.
4 Conclusion la portée philosophique de la théorie des probabilités, par Émile BOREL.

LES PRINCIPES
DE LA THÉORIE DES PROBABILITÉS

ERREURS

ET

MOINDRES CARRÉS

OUVRAGES DU MÊME AUTEUR

LIBRAIRIE GAUTHIER-VILLARS

Probabilités géométriques (tome II du fascicule II du présent Traité, 1926)

LIBRAIRIE ARMAND COLIN

Probabilités, Erreurs (en collaboration avec M. Émile BOREL, 3e édition 1929)

Éléments de Calcul différentiel et intégral (deux volumes, en collaboration avec M. Th. LECONTE, 2e édition, 1929).

COURS DE LA FACULTÉ DES SCIENCES DE TOULOUSE

Cours de Mathématiques générales (2 volumes, *Librairie Édouard Privat, Toulouse* 3e édition, 1930)

TRAITÉ du CALCUL des PROBABILITÉS et de ses APPLICATIONS

Par Émile BOREL

AVEC LA COLLABORATION DE C.-V.-L. CHARLIER, R. DELTHEIL, P. DUBREIL, M. FRÉCHET, H. GALBRUN, J. HAAG, R. LAGRANGE, F. PERRIN, CH. RISSER, P. TRAYNARD

TOME I

LES PRINCIPES DE LA THÉORIE DES PROBABILITÉS

FASCICULE II

ERREURS ET MOINDRES CARRÉS

PAR

R. DELTHEIL

Professeur à la Faculté des Sciences de Toulouse

PARIS

GAUTHIER-VILLARS ET C^ie^, ÉDITEURS

LIBRAIRES DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE

55, Quai des Grands-Augustins, 55

1930

PREFACE

On trouvera dans le présent fascicule un essai d'exposé didactique des théories relatives au problème des erreurs d'observation et a la méthode des moindres carrés

Une première Partie est consacrée à l'étude de *notions préliminaires* comme celle de probabilité des causes et celle de loi de probabilité à une variable, et a l'*énoncé général* des problèmes posés par l'étude des erreurs et la combinaison des observations. Une deuxième Partie traite spécialement de la *loi de Gauss*, en envisageant successivement ses propriétés générales, sa justification théorique, et ses principales applications Enfin une troisième Partie est consacrée à l'étude théorique et pratique de la *méthode des moindres carrés*, conformément aux principes indépendants de toute loi précise de probabilité des erreurs sur lesquels s'appuie ce qu'on appelle souvent la *deuxième théorie de Gauss*

R DELTHEIL

Toulouse, le 11 novembre 1929

ERREURS

ET

MOINDRES CARRÉS

PREMIÈRE PARTIE.

PRINCIPES GÉNÉRAUX.

CHAPITRE I.

PROBABILITES DES CAUSES

1 **Formule de Bayes** — Plusieurs des problèmes poses par la théorie des erreurs et la combinaison des observations rentrent dans le cadre de la théorie des *probabilités des causes*, dont le probleme principal peut s'enoncer d'une manière générale ainsi qu'il suit :

Soit un événement E, *pouvant être réalisé de plusieurs manières s'excluant mutuellement; sachant que l'événement s'est produit, on demande la probabilité pour que ce soit précisément dans l'une, bien déterminée, des hypothèses possibles.*

L'expression de « probabilités des causes » n'est peut-être pas très heureuse, car la notion de *cause* y intervient d'une façon abusive, et la dénomination de « probabilités des hypothèses », proposée par M Fréchet, est certainement préferable; nous nous conformerons cependant à la terminologie consacrée par l'usage.

Chacune des causes a une probabilité déterminée d'intervention *a priori*, et, dans le cas où elle intervient, il lui correspond une

probabilité également déterminée de réalisation de l'événement E; soient $\varpi_1, \varpi_2, \dots, \varpi_n$ les probabilités d'action *a priori* des causes C_1, $C_2, \dots, C_n$, et soient $p_1, p_2, \dots, p_n$ les probabilités correspondantes de réalisation de l'événement.

Évaluons de deux manières la probabilité pour que E soit réalisé par la cause C_i. D'une part, cette probabilité est $\varpi_i p_i$, d'après le principe des probabilités composées. D'autre part, elle est le produit de la probabilité pure et simple de E par la probabilité pour que, E s'étant produit, ce soit précisément par la cause C_i. Cette dernière est la probabilité cherchée P_i, et la probabilité pure et simple de l'événement est

$$\varpi_1 p_1 + \varpi_2 p_2 + \dots + \varpi_n p_n$$

d'après le principe des probabilités totales. La comparaison des deux résultats nous donne la formule

$$P_i = \frac{\varpi_i p_i}{\sum_1^n \varpi_i p_i},$$

appelée *formule de Bayes*.

Cette formule résout le problème envisagé, et son application à des cas concrets ne souffre pas de difficulté lorsque les probabilités *a priori* ϖ_i sont bien déterminées; dans le cas contraire, le problème est mal posé, et n'admet pas de solution certaine.

2. **Étude d'un exemple.** — Examinons l'exemple suivant :

Deux urnes contiennent respectivement A boules, dont a sont blanches, et B boules, dont b sont blanches. Ayant effectué N extractions d'une boule hors de l'une des urnes choisie au hasard, en remettant chaque fois la boule dans l'urne, on constate que k de ces extractions ont donné une boule blanche. Quelle est la probabilité pour que l'urne choisie soit la première ?

La probabilité de réalisation du résultat à partir de la première urne est

$$p_1 = C_N^k \left(\frac{a}{A}\right)^k \left(\frac{A-a}{A}\right)^{N-k},$$

et, à partir de la deuxième,

$$p_2 = C_N^k \left(\frac{b}{B}\right)^k \left(\frac{B-b}{B}\right)^{N-k}.$$

Les probabilités *a priori* étant supposées égales, la probabilité cherchée est

$$P_1 = \frac{p_1}{p_1 + p_2},$$

et nous voyons que

$$\frac{1 - P_1}{P_1} = \frac{p_2}{p_1} = \left(\frac{\beta}{\alpha}\right)^k \left(\frac{1-\beta}{1-\alpha}\right)^{N-k},$$

en posant

$$\frac{a}{A} = \alpha, \qquad \frac{b}{B} = \beta.$$

Examinons les conclusions que l'on peut tirer de ce résultat en supposant que le nombre N des extractions est *très grand*.

Si nous désignons par λ la fréquence $\frac{k}{N}$ d'extraction des boules blanches, et par ω le rapport

$$\frac{\beta^{\lambda}(1-\beta)^{1-\lambda}}{\alpha^{\lambda}(1-\alpha)^{1-\lambda}},$$

nous avons

$$\frac{1 - P_1}{P_1} = \omega^N,$$

de sorte que, pour ω donné, ce rapport tend vers zéro ou vers l'infini avec N, suivant que ω est inférieur ou supérieur à l'unité ; il est égal à 1 quel que soit N si $\omega = 1$.

La fonction

$$y = x^{\lambda}(1-x)^{1-\lambda},$$

dont la dérivée logarithmique s'écrit

$$\frac{y'}{y} = \frac{\lambda}{x} - \frac{1-\lambda}{1-x} = \frac{\lambda - x}{x(1-x)},$$

passe par un maximum pour $x = \lambda$, valeur comprise entre 0 et 1. Si donc on a

$$\alpha < \beta \leqq \lambda \qquad \text{ou} \qquad \alpha > \beta \geqq \lambda,$$

ω est supérieur à 1 et P_1 tend vers zéro ; si, au contraire, on a

$$\beta < \alpha \leqq \lambda \qquad \text{ou} \qquad \beta > \alpha \geqq \lambda$$

ω est inférieur à 1 et P_1 tend vers 1. Si enfin la fréquence constatée λ est comprise entre α et β, il y a lieu, pour pouvoir conclure, de faire directement la comparaison de ω avec l'unité.

3. Importance des probabilités *a priori*. — L'importance des probabilités *a priori* se manifeste d'une manière très nette à propos des problèmes tels que le suivant, relatif à des tirages effectués dans une urne de composition inconnue.

Une urne contient 2N *boules, blanches ou noires, on effectue* 2*n tirages en remettant chaque fois la boule tirée, et l'on constate qu'on a tiré n fois une boule blanche, et n fois une boule noire. On demande la probabilité pour que l'urne contienne* N *boules blanches et* N *boules noires.*

S'il y a A boules blanches et 2N — A noires, la probabilité de réalisation de l'événement observé est

$$p_A = C_{2n}^n \left(\frac{A}{2N}\right)^n \left(\frac{2N-A}{2N}\right)^n.$$

Désignons par ϖ_A la probabilité *a priori* de la composition

$$(A, 2N-A),$$

la probabilité P_N demandée a pour expression

$$P_N = \frac{p_N \varpi_N}{\Sigma_1^{2N-1} p_A \varpi_A},$$

c'est-à-dire

$$P_N = \frac{N^{2n} \varpi_N}{1^n(2N-1)^n \varpi_1 + \ldots + A^n(2N-A)^n \varpi_A + \ldots + (2N-1)^n 1^n \varpi_{2N-1}}.$$

Il n'est possible d'aller plus loin que si l'on fait des hypothèses sur la manière dont a été constituée l'urne envisagée. Si le nombre A a été tiré au sort par l'extraction d'une boule hors d'une urne qui en contenait 2N + 1 numérotées de 0 à 2N, tous les ϖ_i sont égaux, et

$$P_N = \frac{N^{2n}}{N^{2n} + 2[1^n(2N-1)^n + \ldots + (N-1)^n(N+1)^n]}.$$

Par exemple, en admettant que 6 tirages faits dans une urne contenant 8 boules ont donné 3 boules blanches et 3 boules noires, la probabilité pour que l'urne contienne précisément 4 boules blanches et 4 boules noires est

$$\frac{4096}{4096 + 2(343 + 1728 + 3375)} = \frac{1024}{3747},$$

ou environ 0,27; la probabilité *a priori* était seulement $\frac{1}{9}$ ou environ 0,11.

Si la couleur de chacune des 2N boules a été tirée au sort par un coup de pile ou face, les probabilités ϖ_k sont proportionnelles aux coefficients du binome $(x+y)^{2N}$, et l'on a

$$P_N = \frac{C_{2N}^N N^{2n}}{C_{2N}^N N^{2n} + 2[2N\, 1^n (2N-1)^n + \dots + C_{2N}^{N-1}(N-1)^n (N+1)^n]}.$$

Cette deuxième valeur est supérieure à la première, car ϖ_N est la plus grande parmi les probabilités *a priori*, dans le cas numérique $N=4$, $n=3$ examiné plus haut, nous aurions dans la deuxième hypothèse

$$\varpi_4 = \frac{70}{256},$$

soit environ 0,27, et

$$P_4 = \frac{70 \times 4096}{70 \times 4096 + 2[8 \times 343 + 28 \times 1728 + 56 \times 3375]} = \frac{280}{749},$$

soit environ 0,37.

4. **Cas des probabilités continues.** — Étudions d'abord le problème simple suivant, dont l'énoncé suggestif est dû à M. Émile Borel

Problème. — *On lance en l'air un solide polyédrique dont certaines faces sont peintes en blanc et les autres en noir, on ne sait rien sur le nombre ni les dimensions des faces, mais on sait que $p+q$ expériences ont donné pour résultats de laisser retomber le solide p fois sur une face blanche et q fois sur une face noire. Quelle idée peut-on se faire, d'après ce renseignement, de la probabilité pour que le solide retombe sur une face blanche?*

Si x est cette probabilité inconnue, la probabilité correspondante de réalisation de l'événement observé a pour expression

$$\frac{(p+q)!}{p!\,q!} x^p (1-x)^q$$

Soit $\varphi(x)\,dx$ la probabilité élémentaire *a priori* d'intervention de la cause d'après laquelle la probabilité pour que le solide retombe sur une face blanche est comprise entre x et $x+dx$. La formule de

Bayes donne, pour la probabilité *a posteriori* d'intervention de cette cause, le résultat des essais signalés étant connu, l'expression

$$P\,dx = \frac{x^p(1-x)^q\varphi(x)\,dx}{\int_0^1 x^p(1-x)^q\varphi(x)\,dx},$$

en supprimant le facteur combinatoire qui figure à la fois au numérateur et au dénominateur.

Si l'on ne sait rien sur $\varphi(x)$, on est fondé *à considérer toutes les valeurs de x comme également probables a priori*, ce qui revient à faire $\varphi(x) = 1$. C'est cette hypothèse que l'on nomme souvent *hypothèse de Bayes*; si naturelle qu'elle paraisse au point de vue du bon sens, elle n'est nullement imposée par la logique. Les problèmes de la nature actuellement envisagée sont donc de ceux, nombreux en matière de probabilités continues, dont la solution dépend d'une *fonction positive arbitraire*.

Admettant l'hypothèse de Bayes, nous pouvons calculer la probabilité $P\,dx$; soit $J(p, q)$ l'intégrale

$$\int_0^1 x^p(1-x)^q\,dx$$

qui figure au dénominateur; nous avons en intégrant par parties

$$J(p, q) = \frac{q}{p+1} J(p+1, q-1),$$

et cette relation de récurrence, appliquée à $J(p+1, q-1)$, $J(p+2, q-2)$, ..., $J(p+q, 0)$ donne immédiatement

$$J(p, q) = \frac{p!\,q!}{(p+q+1)!},$$

de sorte que

$$P\,dx = \frac{(p+q+1)!}{p!\,q!} x^p(1-x)^q\,dx.$$

La dérivée logarithmique du produit $x^p(1-x)^q$ étant

$$\frac{p}{x} - \frac{q}{1-x},$$

la valeur *la plus probable* de x est

$$X = \frac{p}{p+q},$$

fréquence observée pour l'arrivée des faces blanches, la *valeur probable* est

$$\bar{x} = \int_0^1 xP\,dx = \frac{\int_0^1 x^{p+1}(1-x)^q\,dx}{\int_0^1 x^p(1-x)^q\,dx},$$

ou

$$\bar{x} = \frac{J(p+1, q)}{J(p, q)} = \frac{p+1}{p+q+2},$$

expression très peu différente de X si p et q sont très grands.

L'écart entre la valeur éventuelle x et la valeur probable $\bar{x}$ est

$$\xi = x - \bar{x};$$

son carré a pour valeur probable

$$\overline{\xi^2} = \int_0^1 P(x-\bar{x})^2\,dx,$$

c'est-à-dire

$$\int_0^1 Px^2\,dx - 2\bar{x}\int_0^1 Px\,dx + \bar{x}^2\int_0^1 P\,dx;$$

comme nous avons

$$\int_0^1 P\,dx = 1, \qquad \int_0^1 Px\,dx = \bar{x},$$

$$\int_0^1 Px^2\,dx = \frac{(p+1)(p+2)}{(p+q+2)(p+q+3)},$$

il vient en définitive

$$\overline{\xi^2} = \frac{(p+1)(p+2)}{(p+q+2)(p+q+3)} - \frac{(p+1)^2}{(p+q+2)^2}$$

$$= \frac{(p+1)(q+1)}{(p+q+2)^2(p+q+3)}.$$

Si l'on change p en np, q en nq, ce carré moyen est, pour np et nq grands, sensiblement divisé par n; on voit que l'écart quadra-

tique moyen $\sqrt{\overline{\xi^2}}$ est sensiblement proportionnel à *l'inverse de la racine carrée du nombre des expériences.*

5. Étude d'un problème plus général — Dans le problème précédent, les causes forment un ensemble infini et continu, mais les résultats possibles des $p+q$ expériences sont en nombre fini. Les questions que nous aurons à étudier dans ce volume et qui relèvent de la théorie des probabilités des causes se rattachent à un problème plus général, dans l'énoncé duquel les causes et les résultats forment deux ensembles infinis et continus.

Soit x une variable, susceptible de prendre toutes les valeurs d'un intervalle (a, b), la probabilité élémentaire donnée *a priori* étant

$$\varpi\, dx = \varphi(x)\, dx,$$

soit, d'autre part, y une nouvelle variable, susceptible de prendre, pour chaque valeur de x prise dans (a, b), des valeurs éventuelles d'un intervalle (α, β), les bornes de cet intervalle et la probabilité élémentaire

$$p\, dy = f(x, y)\, dy$$

dépendant de x. *L'expérience ayant fourni une valeur y_1 de cette variable, on demande la nouvelle probabilité élémentaire* $\mathrm{P}\, dx$ *résultant, pour la valeur de x, de ce renseignement.*

D'après la formule de Bayes, nous avons

$$\mathrm{P}\, dx = \frac{\varphi(x)\, dx\, f(x, y_1)\, dy_1}{\displaystyle\int_a^b \varphi(x)\, dx\, f(x, y_1)\, dy_1},$$

ou, en supprimant dy_1 qui est en facteur haut et bas,

$$\mathrm{P}\, dx = \frac{\varphi(x) f(x, y_1)\, dx}{\displaystyle\int_a^b \varphi(x) f(x, y_1)\, dx}.$$

Plusieurs expériences ayant donné pour y les valeurs $y_1, y_2, \ldots, y_n$, on peut se poser la même question. Le numérateur de la formule de Bayes est alors

$$\varphi(x)\, dx\, f(x, y_1)\, dy_1 \ldots f(x, y_n)\, dy_n;$$

les différentielles dy_1, dy_2, ..., dy_n sont en facteur haut et bas, et il reste

$$P\,dx = \frac{\varphi(x) f(x, y_1) \cdots f(x, y_n)\,dx}{\int_a^b \varphi(x) f(x, y_1) \cdots f(x, y_n)\,dx}.$$

Plus généralement, soient x, y, z trois variables, coordonnées d'un point M de l'espace, susceptible *a priori* de prendre les diverses positions d'un certain domaine D, la probabilité élémentaire étant

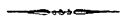
$$\varpi\,dx\,dy\,dz = \varphi(x, y, z)\,dx\,dy\,dz.$$

Soient, d'autre part, u, v deux nouvelles variables, coordonnées d'un point μ dans un plan, susceptible de prendre des positions éventuelles dans un domaine Δ, le domaine Δ et la probabilité élémentaire

$$p\,du\,dv = f(x, y, z, u, v)\,du\,dv$$

dépendant de x, y, z.

Plusieurs expériences ayant donné pour le point μ les positions (u_1, v_1), ..., (u_n, v_n), la nouvelle valeur résultant de ces renseignements pour la probabilité élémentaire relative au point M est

$$P\,dx\,dy\,dz = \frac{\varphi(x, y, z) f(x, y, z, u_1, v_1) \cdots f(x, y, z, u_n, v_n)\,dx\,dy\,dz}{\int\int\int_D \varphi(x, y, z) f(x, y, z, u_1, v_1) \cdots f(x, y, z, u_n, v_n)\,dx\,dy\,dz}.$$

CHAPITRE II.

LOIS DE PROBABILITÉ A UNE VARIABLE

I — GÉNÉRALITÉS.

6. **Détermination d'une loi de probabilité.** — L'une des notions fondamentales de la théorie mathématique des probabilités est celle de grandeur aléatoire ou *variable éventuelle*. On donne ce nom à une variable x susceptible de prendre diverses valeurs suivant que se présentent diverses éventualités dont les probabilités sont déterminées.

La détermination la plus générale de la *loi de probabilité* suivie par une variable éventuelle x est obtenue par la connaissance de la fonction exprimant, pour chaque valeur de x, la probabilité pour que la variable ait une valeur comprise entre $-\infty$ et x. Cette fonction $F(x)$ a déjà fait l'objet d'une étude importante dans le fascicule I du présent Traité, sous la dénomination, employée en statistique, de *fonction de répartition*. Nous l'appellerons avec M. Paul Lévy la *fonction des probabilités totales*; elle varie de 0 à 1 quand x varie de $-\infty$ à $+\infty$.

Étudions d'abord quelques exemples particuliers.

1° Si la variable éventuelle peut prendre seulement des valeurs en nombre fini $x_1, x_2, \ldots, x_n$, les probabilités correspondantes ayant les valeurs $p_1, p_2, \ldots, p_n$, dont la somme est l'unité, la fonction $F(x)$ reste constante dans chacun des intervalles $(-\infty, x_1)$, (x_1, x_2), ..., $(x_n, +\infty)$; elle admet en chacun des points x_i une discontinuité de première espèce. C'est une *fonction en escalier* comme celles envisagées dans l'étude du fascicule I signalée plus haut (Chap. VI : *Statistique*). Par exemple, si x désigne le nombre de résultats « pile » que l'on peut observer sur une série de quatre coups de pile ou face, les valeurs possibles de la variable sont 0, 1, 2, 3, 4, et leurs probabilités

respectives sont $\frac{1}{16}$, $\frac{4}{16}$, $\frac{6}{16}$, $\frac{4}{16}$, $\frac{1}{16}$; la *courbe des probabilités totales* se composera de la partie négative toute entière de l'axe Ox, du segment $(0, 1)$ de la droite $y = \frac{1}{16}$, du segment $(1, 2)$ de la droite $y = \frac{5}{16}$, du segment $(2, 3)$ de la droite $y = \frac{11}{16}$, du segment $(3, 4)$ de la droite $y = \frac{15}{16}$ et de la partie illimitée de la droite $y = 1$ située à droite de l'ordonnée $x = 4$. La valeur de la fonction $F(x)$ en un point de discontinuité dépend de la convention que l'on fait sur le point de savoir si la valeur x est ou n'est pas incluse dans l'intervalle $(-\infty, x)$, dans la négative, la fonction est discontinue à gauche, continue à droite de chacun des points $x_1, x_2, \ldots, x_n$.

2° Les valeurs que peut prendre x, tout en restant isolées, forment dans certaines applications un *ensemble infini dénombrable*. Si nous les désignons par

$$\ldots, x_{-n}, x_{-n+1}, \ldots, x_{-1}, x_0, x_1, \ldots, x_{n-1}, x_n, \ldots$$

les probabilités respectives étant

$$\ldots, p_{-n}, p_{-n+1}, \ldots, p_{-1}, p_0, p_1, \ldots, p_{n-1}, p_n, \ldots$$

les deux séries positives

$$p_0 + p_1 + p_2 + \ldots + p_n + \ldots$$
$$p_{-1} + p_{-2} + \ldots + p_{-n} + \ldots$$

doivent être convergentes et le total de leurs deux sommes doit être l'unité.

3° Les valeurs possibles de x peuvent être toutes les valeurs d'un intervalle (a, b), ou même de l'intervalle $(-\infty, +\infty)$, la probabilité pour que la variable prenne une valeur comprise entre x et $x + dx$ étant la *probabilité élémentaire* $f(x)\,dx$. La loi de probabilité est alors dite *absolument continue*, et la fonction sommable $f(x)$ est appelée par M. Paul Lévy la *fonction des probabilités élémentaires;* la courbe $y = f(x)$ fournit, par les aires qu'elle délimite, une représentation extrêmement importante de la loi de probabilité.

Dans le cas général, on peut, ainsi qu'il est montré dans le Chapitre VI du fascicule I du présent Traité, considérer $F(x)$ comme

définissant une *distribution de masse* le long de l'axe Ox. Si $F(x)$ est une fonction en escalier, ces masses sont finies et concentrées en certains points formant un ensemble fini ou dénombrable. Si la loi de probabilité est définie par une fonction des probabilités élémentaires, les masses sont réparties d'une manière continue le long de Ox, avec une densité représentée par la fonction positive sommable $f(x)$. Mais il peut se présenter aussi (*voir* fascicule I, p. 111 à 119) des masses réparties avec une densité infinie en des points formant un ensemble de mesure nulle, sans qu'il y ait aucune masse finie concentrée en un point.

Dans une étude générale, il y a donc lieu de prévoir l'existence de masses des *trois catégories* ainsi mises en évidence, cependant il nous arrivera d'examiner, à propos de certaines questions, les cas simples où il y a seulement des masses isolées et où il y a seulement une distribution continue de densité sommable.

7. Valeurs moyennes et moments. — Étant donnée une fonction $g(x)$ de la variable éventuelle x susceptible de prendre les valeurs $x_1, x_2, \ldots, x_n$ avec les probabilités respectives $p_1, p_2, \ldots, p_n$, nous savons qu'on appelle *valeur probable* ou *valeur moyenne* de cette fonction, l'expression

$$\overline{g(x)} = p_1 g(x_1) + p_2 g(x_2) + \ldots + p_n g(x_n).$$

Cette valeur moyenne se présente comme la somme d'une série si les valeurs x_i forment un ensemble infini dénombrable, et comme la valeur de l'intégrale

$$\overline{g(x)} = \int_{-\infty}^{+\infty} g(x) f(x)\, dx,$$

s'il y a une fonction des probabilités élémentaires $f(x)$.

Pour définir et représenter cette valeur moyenne dans le cas général, supposons la fonction $g(x)$ continue, et plaçons-nous d'abord dans le cas où les valeurs éventuelles de x sont celles d'un intervalle fini (a, b). Si nous divisons cet intervalle en intervalles partiels par les points de division $x_1, x_2, \ldots, x_{n-1}$, la probabilité relative à l'intervalle (x_{i-1}, x_i) est $F(x_i) - F(x_{i-1})$, donc en désignant par ξ_i une valeur de cet intervalle, la portion correspondante de la valeur moyenne de $f(x)$ est voisine de

$$g(\xi_i)[F(x_i) - F(x_{i-1})].$$

On démontre en Analyse qu'une somme de quantités de cette nature admet une limite indépendante du choix des ξ_i lorsque n tend vers l'infini, tous les intervalles partiels tendant vers zéro, pourvu que la fonction $g(x)$ soit continue, et que $F(x)$ soit à variation bornée, ce qui est assurément le cas actuellement; cette limite est *l'intégrale de Stieltjes*

$$\int_a^b g(x)\,dF(x)$$

dont l'extension à l'intervalle $(-\infty, +\infty)$ se fait comme celle de l'intégrale définie usuelle.

Nous définissons donc la valeur moyenne de $g(x)$ dans le cas général par l'intégrale de Stieltjes

$$\overline{g(x)} = \int_{-\infty}^{+\infty} g(x)\,dF(x)$$

qui comprend comme cas particuliers les deux expressions de $\overline{g(x)}$ relatives aux cas des masses isolées et d'une distribution de masse absolument continue.

On appelle *moments* de la distribution de masse associée à une loi de probabilité *les valeurs moyennes des puissances entières positives de la variable*, le moment d'ordre h est

$$\mu_h = \int_{-\infty}^{+\infty} x^h\,dF(x),$$

ce moment est nul pour *h impair* si la distribution de masse est *symétrique* par rapport à l'origine (par exemple s'il y a une fonction paire des probabilités élémentaires); alors on a souvent à considérer les valeurs moyennes des puissances impaires de $|x|$, c'est-à-dire les quantités

$$\int_{-\infty}^{+\infty} |x|^{2k+1}\,dF(x)$$

8. **L'écart quadratique moyen; son importance** — Le moment d'ordre 1 est la valeur moyenne $\mu_1 = \xi$ de la variable; il est nul dans le cas d'une loi de probabilité symétrique. On emploie souvent, dans le calcul des probabilités, la différence $x - \xi$ entre la variable et sa valeur moyenne, et que l'on nomme *l'écart*.

La valeur moyenne du carré de l'écart

$$m^2 = \int_{-\infty}^{+\infty} (x-\xi)^2 \, dF(x)$$

a pour expression

$$\int_{-\infty}^{+\infty} x^2 \, dF(x) - 2\xi \int_{-\infty}^{+\infty} x \, dF(x) + \xi^2 \int_{-\infty}^{+\infty} dF(x)$$

ou, les trois intégrales de la somme ci-dessus ayant pour valeurs respectives μ_2, $\mu_1 = \xi$ et 1,

$$m^2 = \mu_2 - \mu_1^2.$$

On donne souvent à m le nom d'*écart quadratique moyen*, c'est là une quantité d'une importance capitale, dont la valeur caractérise l'ordre de grandeur des écarts d'une manière commode et remarquable, que l'on peut préciser ainsi qu'il suit, au moyen du *théorème de Bienaymé-Tchebychef*. *Quelle que soit la loi de probabilité, la probabilité pour que l'écart soit supérieur ou égal en valeur absolue à αm est inférieure ou égale à $\frac{1}{\alpha^2}$.*

Considérons en effet la somme

$$\int_{-\infty}^{\xi - m\alpha} (x-\xi)^2 \, dF(x) + \int_{\xi + m\alpha}^{+\infty} (x-\xi)^2 \, dF(x),$$

qui ne représente qu'une fraction de l'intégrale définissant m^2, et par conséquent est inférieure ou égale à m^2. Puisque, dans les deux intervalles d'intégration, $(x-\xi)^2$ est supérieur à $m^2\alpha^2$, cette somme est supérieure ou égale à

$$m^2\alpha^2 \left[\int_{-\infty}^{\xi - m\alpha} dF(x) + \int_{\xi + m\alpha}^{+\infty} dF(x) \right],$$

produit de $m^2\alpha^2$ par la probabilité considérée. Celle-ci est donc bien inférieure ou égale à $\frac{1}{\alpha^2}$.

On peut énoncer un théorème analogue pour toute puissance paire de l'écart, mais le rôle privilégié de la deuxième puissance se justifie en raison du fait qu'elle est la plus simple et que son emploi conduit à des résultats d'une grande fécondité, dont quelques-uns feront l'objet de plusieurs Chapitres du présent fascicule.

II — LE PROBLÈME DES MOMENTS [1]

9. Le problème algébrique d'ordre n. — Le problème général des moments consiste dans la détermination d'une loi de probabilité par la connaissance de la suite illimitée de ses moments.

Soit donnée une fonction des probabilités totales $F(x)$ que nous supposons *ne pas se réduire à une fonction à un nombre fini d'escaliers*, désignons par

$$c_0, \quad c_1, \quad c_2, \quad \ldots, \quad c_{2n-1}, \quad \ldots$$

la suite illimitée des moments qui lui correspondent.

Un premier problème, envisagé par Tchebychef et que nous appellerons avec M. Castelnuovo *problème algébrique d'ordre n*, consiste à déterminer une fonction en escalier définie par les masses positives

$$a_1, \quad a_2, \quad \ldots, \quad a_n$$

disposées aux points de discontinuité

$$x_1, \quad x_2, \quad \ldots, \quad x_n$$

par les conditions que les $2n$ premiers moments de cette fonction soient précisément $c_0, c_1, c_2, \ldots, c_{2n-1}$.

Ces conditions s'expriment par les relations

$$(1) \quad \begin{cases} a_1 + a_2 + \ldots + a_n = c_0, \\ a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = c_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_1 x_1^{2n-1} + a_2 x_2^{2n-1} + \ldots + a_n x_n^{2n-1} = c_{2n-1}. \end{cases}$$

Proposons-nous de former l'équation

$$H(x) = \lambda_0 + \lambda_1 x + \ldots + \lambda_n x^n = 0$$

dont $x_1, x_2, \ldots, x_n$ sont les racines. Nous avons, quel que

(1) Cette étude fait en partie double emploi avec celle du Chapitre VI du fascicule I, j'ai cru devoir la maintenir ici pour son adaptation au but poursuivi. (*Voir* plus loin, Chap. VI du présent volume.)

soit $i = 1, 2, \ldots, n$, la relation $H(x_i) = 0$; formons les sommes

$$\sum_{i=1}^{i=n} a_i x_i^k H(x_i)$$

pour $k = 0, 1, 2, \ldots, n-1$, ces sommes sont nulles, et nous avons par conséquent, d'après (1), les relations

$$(2) \qquad \begin{cases} \lambda_0 c_0 + \lambda_1 c_1 + \ldots + \lambda_n c_n = 0, \\ \lambda_0 c_1 + \lambda_1 c_2 + \ldots + \lambda_n c_{n+1} = 0, \\ \ldots \ldots \ldots \ldots \\ \lambda_0 c_{n-1} + \lambda_1 c_n + \ldots + \lambda_n c_{2n-1} = 0. \end{cases}$$

Il s'ensuit que $\lambda_0, \lambda_1, \ldots, \lambda_n$ *sont proportionnels aux déterminants d'ordre n tirés du tableau*

$$\begin{Vmatrix} c_0 & c_1 & \ldots & c_n \\ \ldots & \ldots & \ldots & \ldots \\ c_{n-2} & c_{n-1} & \ldots & c_{2n-2} \\ c_{n-1} & c_n & \ldots & c_{2n-1} \end{Vmatrix},$$

ce qui détermine le polynome $H(x)$ à un facteur constant près.

Il importe de s'assurer que le mineur correspondant au terme en x^n dans ce polynome n'est pas nul; ce mineur s'écrit

$$\Delta_n = \begin{vmatrix} c_0 & c_1 & \ldots & c_{n-1} \\ c_1 & c_2 & \ldots & c_n \\ \ldots & \ldots & \ldots & \ldots \\ c_{n-2} & c_{n-1} & \ldots & c_{2n-2} \end{vmatrix},$$

donc, en remplaçant les moments par leurs valeurs exprimées au moyen de coordonnées $y_1, y_2, \ldots, y_n$ d'un espace à n dimensions, Δ_n se présente comme l'intégrale, étendue à tout cet espace, de l'élément différentiel

$$\begin{vmatrix} 1 & y_1 & \ldots & y_1^{n-1} \\ y_2 & y_2^2 & \ldots & y_2^n \\ \ldots & \ldots & \ldots & \ldots \\ y_n^{n-1} & y_n^n & \ldots & y_n^{2n-1} \end{vmatrix} dF(y_1)\,dF(y_2)\ldots dF(y_n),$$

produit du déterminant de Van der Monde

$$\delta = \begin{vmatrix} 1 & y_1 & \dots & y_1^{n-1} \\ 1 & y_2 & \dots & y_2^{n-1} \\ \cdot & \cdot & \dots & \cdot \\ 1 & y_n & \dots & y_n^{n-1} \end{vmatrix}$$

par $y_2 y_3^2 \dots y_n^{n-1}\, dF(y_1)\, dF(y_2) \dots dF(y_n)$.

En permutant de toutes les manières possibles, au nombre de $n!$, les variables $y_1, y_2, \dots, y_n$, on met l'élément différentiel ci-dessus sous forme du produit de $\delta\, dF(y_1)\, dF(y_2) \dots dF(y_n)$ par les $n!$ termes successifs, chacun avec son signe, dont la somme est δ. Il en résulte que $n!\Delta_n$ est l'intégrale, étendue à tout l'espace, de l'élément différentiel

$$d\Omega = \delta^2\, dF(y_1)\, dF(y_2) \dots dF(y_n)$$

essentiellement positif ou nul. Cet élément ne peut être nul dans tout l'espace que si l'une des quantités $y_i - y_j$ ou l'une des quantités $dF(y_k)$ est nulle en tout point de l'espace, ce qui exige que $F(x)$ *se réduise à une fonction à moins de n escaliers*. S'il y a, en effet, p escaliers correspondant aux points de discontinuité ξ_1, ξ_2, ..., ξ_p, tous les $dF(y_k)$ sont nuls à moins que chacun des y_k ne soit égal à l'un des ξ_i; alors si le nombre p des ξ_i est inférieur à n, il y a des y_k égaux entre eux, et δ est nul.

Il résulte de cette étude que, moyennant les hypothèses faites, les abscisses x_1, x_2, ..., x_n sont les zéros d'un polynome $Q_n(x)$ que nous savons former. Nous l'écrirons sous la forme

$$Q_n(x) = \frac{(-1)^n}{\Delta_n} \begin{vmatrix} 1 & x & x^2 & \dots & x^n \\ c_0 & c_1 & c_2 & \dots & c_n \\ \cdot & \cdot & \cdot & \dots & \cdot \\ c_{n-1} & c_n & c_{n+1} & \dots & c_{2n-1} \end{vmatrix},$$

de manière à donner au terme en x^n le coefficient 1.

10. Propriétés des polynomes $Q_n(x)$. — Nous allons montrer que, quel que soit n, le polynome $Q_n(x)$ *a n racines réelles et distinctes*, et qu'il leur correspond *des masses a_1, a_2, ..., a_n positives*. Ces propriétés sont vraies pour $n = 1$, car alors, d'après ce qui précède,

$$Q_1(x) = x - \frac{c_1}{c_0},$$

donc

$$r_1 = \frac{c_1}{c_0},$$

et $a_1 = c_0$ qui est égal à 1 si $F(x)$ est bien une fonction des probabilités totales. Nous allons les étendre au cas général

Pour cela, notons que les relations (2) peuvent s'écrire

$$\int_{-\infty}^{+\infty} x^k Q_n(x)\, dF(x) = 0,$$

où $k = 0, 1, 2, \ldots, n-1$. On peut donc dire que *le polynome $Q_n(x)$ est complètement déterminé par la condition que l'intégrale*

$$\int_{-\infty}^{+\infty} P Q_n\, dF(x)$$

soit nulle, P *étant un polynome* ARBITRAIRE *de degré $n-1$ au plus*, et que, d'autre part, le coefficient de x_n dans $Q_n(x)$ soit égal à 1 (la première condition déterminant à elle seule le polynome à un facteur constant près)

La division de Q_n par Q_{n-1} étant représentee par l'identite

$$Q_n(x) = (x - \alpha_{n-1}) Q_{n-1}(x) + R(x)$$

il est clair que le reste $R(x)$ possède *la propriété d'orthogonalité* ci-dessus vis-à-vis de tout polynome $P(x)$ de degré $n-3$ au plus, et par conséquent est identique à $Q_{n-2}(x)$ a un certain facteur constant près; nous avons donc l'identite

$$Q_n(x) = (x - \alpha_{n-1}) Q_{n-1}(x) - \omega_{n-1} Q_{n-2}(x)$$

Pour calculer la constante ω_{n-1}, multiplions par $Q_{n-2}\, dF(x)$ les deux membres de cette identité, et intégrons de $-\infty$ à $+\infty$; nous aurons zéro au premier membre, et, au second membre, l'integrale coefficient de α_{n-1} sera nulle aussi, il restera

$$\int_{-\infty}^{+\infty} x Q_{n-1} Q_{n-2}\, dF(x) = \omega_{n-1} \int_{-\infty}^{+\infty} Q_{n-2}^2\, dF(x),$$

d'où l'on tire, en remarquant que l'intégrale du premier membre est égale à

$$\int_{-\infty}^{+\infty} Q_{n-1}^2\, dF(x),$$

parce que le produit $x Q_{n-2}$ ne diffère de Q_{n-1} que par un polynome de degré $n-2$ au plus,

$$\omega_{n-1} = \frac{\int_{-\infty}^{+\infty} Q_{n-1}^2 \, dF(x)}{\int_{-\infty}^{+\infty} Q_{n-2}^2 \, dF(x)}.$$

Le facteur ω_{n-1} est donc positif Par conséquent Q_n et Q_{n-2} prennent des valeurs de signes contraires lorsque x est égal à une racine de Q_{n-1}. Il en résulte, en considérant de proche en proche, les polynomes

$$Q_0(x) = 1 \qquad Q_1(x) = x - \frac{c_1}{c_0}, \qquad Q_2(x),$$

que $Q_n(x)$ admet n racines reelles, distinctes et séparées par les racines de $Q_{n-1}(x)$ Les polynomes $Q_n(x)$ forment une *suite de Sturm* généralisee

Montrons maintenant que tous les a_i sont positifs Les conditions (1) entraînent l'identité

$$(3) \qquad \sum_{i=1}^{i=n} a_i \theta(x_i) = \int_{-\infty}^{+\infty} \theta(x) \, dF(x)$$

pour tout polynome $\theta(x)$ de degré $2n-1$ au plus

Prenons

$$\theta(x) = \frac{Q_n^2(x)}{(x-x_i)^2},$$

dans le premier membre de l'identité (3), tous les termes sont nuls, sauf le terme $a_i \theta(x_i)$ qu'on peut écrire $a_i Q_n'^2(x_i)$, l'identité donne alors pour a_i la valeur

$$a_i = \frac{1}{Q_n'^2(x_i)} \int_{-\infty}^{+\infty} (x-x_1)^2 \dots (x-x_{i-1})^2 (x-x_{i+1})^2 \dots (x-x_n)^2 \, dF(x),$$

qui est bien positive.

11. Probleme général des moments. — Ce problème, énoncé au § 9, est la forme limite du problème algébrique d'ordre n lorsque n augmente indéfiniment.

Nous avons établi que, si $F(x)$ ne se réduit pas à une fonction à un nombre fini d'escaliers, le problème algébrique d'ordre n conduit à une fonction des probabilités totales $G_n(x)$, à n escaliers, dont les $2n$ premiers moments sont ceux de $F(x)$.

Tchebychef a démontré deux inégalités fondamentales d'où il résulte que *toutes les marches de la courbe en escalier* $y = G_n(x)$ *traversent la courbe* $y = F(x)$.

Appliquons l'identité (3) ci-dessus au polynome $\theta = f_{n-1}(x)$, de degré $n - 1$, déterminé par les conditions

$$f_{n-1}(x_1) = f_{n-1}(x_2) = \ldots = f_{n-1}(x_k) = 1,$$
$$f_{n-1}(x_{k+1}) = f_{n-1}(x_{k+2}) = \ldots = f_{n-1}(x_n) = 0.$$

L'identité devient

$$a_1 + a_2 + \ldots + a_k = \int_{-\infty}^{+\infty} f_{n-1}(x)\, dF(x),$$

et, d'après la propriété fondamentale des polynomes $Q_n(x)$, nous pouvons remplacer $f_{n-1}(x)$, au second membre de cette relation, par

$$\Phi(x) = f_{n-1}(x) + \varphi_{n-2}(x) Q_n(x),$$

où $\varphi_{n-2}(x)$ désigne un polynome arbitraire de degré $n - 2$.

Notons que, quel que soit ce polynome, $\Phi(x)$ est égal à 1 pour $x = x_1, x_2, \ldots, x_k$ et à 0 pour $x = x_{k+1}, \ldots, x_n$, si nous déterminons maintenant $\varphi_{n-2}(x)$ de manière que $x_1, x_2, \ldots, x_{k-1}$ et $x_{k+1}, \ldots, x_n$ soient racines de $\Phi'(x)$, nous avons les $n - 1$ conditions

$$f'_{n-1}(x_i) + \varphi_{n-2}(x_i) Q'_n(x_i) = 0,$$

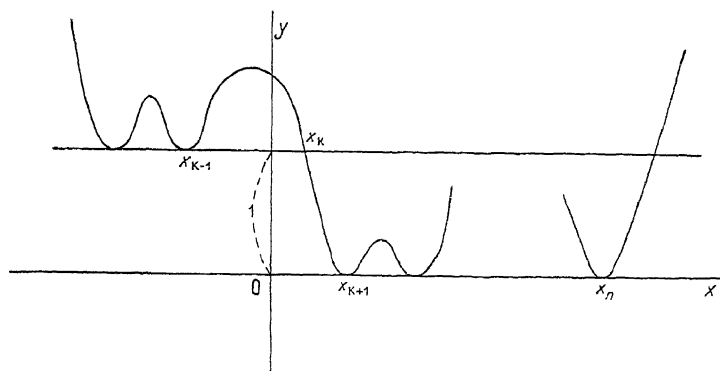
pour $i = 1, 2, \ldots, k - 1, k + 1, \ldots, n$. Ces conditions déterminent sans ambiguïté le polynome $\Phi(x)$.

Or, si l'on trace (*fig.* 1) la courbe $y = \Phi(x)$, qui touche la droite $y = 1$ aux points $x_1, x_2, \ldots, x_{k-1}$, la coupe au point x_k, et touche Ox aux points $x_{k+1}, \ldots, x_n$, on constate que $\Phi(x)$ est *supérieur à* 1 pour toutes les valeurs de l'intervalle $(-\infty, x_k)$, et *supérieur à* 0 pour toutes celles de l'intervalle $(x_k, +\infty)$.

Le second membre de l'identité

$$a_1 + a_2 + \ldots + a_k = \int_{-\infty}^{+\infty} \Phi(x)\, dF(x)$$

est donc supérieur à l'intégrale $\int_{-\infty}^{x_k} dF(x)$. En adaptant la même

Fig. 1

démonstration à l'intervalle $(x_{k+1}, +\infty)$, on obtient l'inégalité

$$a_{k+1} + a_{k+2} + \dots + a_n > \int_{x_{k+1}}^{+\infty} dF(x).$$

Et, réunissant enfin les deux résultats, compte tenu de la relation

$$a_1 + a_2 + \dots + a_n = \int_{-\infty}^{+\infty} dF(x),$$

on obtient la double inégalité

$$(4) \qquad \int_{-\infty}^{x_k} dF(x) < a_1 + a_2 + \dots + a_k < \int_{-\infty}^{x_{k+1}} dF(x),$$

démontrant que les points M_k et M_{k-1}, d'abscisses x_k et x_{k+1}, et d'ordonnée commune $a_1 + a_2 + \dots + a_k$, sont situés de part et d'autre de la courbe $y = F(x)$ (*fig.* 2).

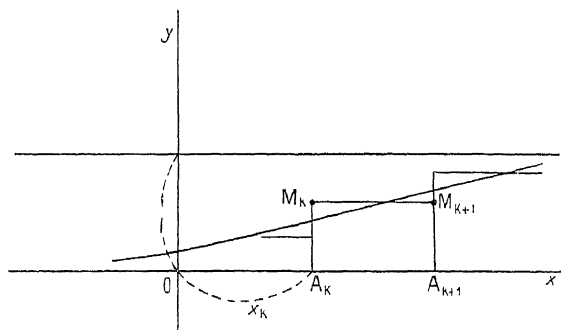
On peut aussi conclure des inégalités de Tchebychef, par l'examen de la figure 2, que *la différence*

$$F(x) - G_n(x)$$

est constamment inférieure en module à la plus grande des deux masses a_k entre lesquelles se trouve placé le point x.

Si l'on fait tendre n vers l'infini, la différence précédente sera tou-

jours inférieure à un nombre donné ε si chacune des masses a_1, a_2, ..., a_n est moindre que ε. Cette condition, *que les masses* a_1, a_2, ..., a'_n *tendent uniformément vers zéro*, est donc une condi-

Fig.

tion suffisante de *convergence uniforme* des fonctions en escalier $G_n(x)$ vers la fonction $F(x)$. Nous prendrons ce résultat comme point de départ pour la justification de la loi des erreurs de Gauss par la méthode des moments (Chap. VI).

III. — FONCTION CARACTÉRISTIQUE D'UNE LOI DE PROBABILITÉ

12. Définition et exemples. — La notion de *fonction caractéristique* est due à Cauchy; elle a été reprise par Poincaré, et les travaux récents de M. Paul Lévy en ont montré la grande fécondité.

Étant donnée une loi de probabilité définie par sa fonction des probabilités totales $F(x)$, la fonction caractéristique, au sens de Cauchy, est la valeur moyenne

$$\varphi(\theta) = \int_{-\infty}^{+\infty} e^{x\theta}\, dF(x)$$

de $e^{x\theta}$; au sens de M. Paul Lévy, c'est la valeur moyenne

$$\Phi(t) = \int_{-\infty}^{+\infty} e^{ixt}\, dF(x)$$

de e^{ixt}. Nous utiliserons les deux définitions.

Formons la fonction caractéristique au sens de Cauchy dans quelques cas particuliers simples.

1° Si $F(x)$ se réduit à une fonction à n escaliers, on dit quelquefois que x est une *variable éventuelle d'ordre n*, on a alors avec les notations du § 6,

$$\varphi(\theta) = p_1 e^{\theta x_1} + p_2 e^{\theta x_2} + \dots + p_n e^{\theta x_n}.$$

Cette fonction est développable en série sous la forme

$$(1) \qquad \varphi(\theta) = \mu_0 + \frac{\mu_1}{1}\theta + \frac{\mu_2}{2!}\theta^2 + \dots + \frac{\mu_h}{h!}\theta^h + \dots$$

les coefficients $\mu_0, \mu_1, \dots, \mu_h$ désignant les *moments*. Il est clair que $\varphi(\theta)$ est solution d'une équation différentielle linéaire et homogène d'ordre n à coefficients constants, de la forme

$$A_0 \varphi + A_1 \varphi' + A_2 \varphi'' + \dots + A_n \varphi^{(n)} = 0$$

dont l'équation caractéristique a pour racines $x_1, x_2, \dots, x_n$; il en résulte, en portant le développement (1) dans cette équation, les relations

$$A_0\mu_0 + A_1\mu_1 + \dots + A_n\mu_n = 0$$
$$A_0\mu_1 + A_1\mu_2 + \dots + A_n\mu_{n+1} = 0$$
$$\dots$$
$$A_0\mu_k + A_1\mu_{k+1} + \dots + A_n\mu_{n+k} = 0,$$

dont les n premières sont, aux notations près, les relations (2) du § 8; et nous voyons que l'équation différentielle vérifiée par $\varphi(\theta)$ peut s'écrire

$$(2) \qquad \begin{vmatrix} \varphi & \varphi' & \varphi'' & \dots & \varphi^{(n)} \\ \mu_0 & \mu_1 & \mu_2 & \dots & \mu_n \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{n-1} & \mu_n & \mu_{n+1} & \dots & \mu_{2n-1} \end{vmatrix} = 0.$$

Par ailleurs, la suite illimitée des moments est entièrement déterminée par la donnée des $2n$ premiers, les moments suivants pouvant se calculer de proche en proche au moyen des relations linéaires écrites plus haut.

2° Imaginons que la variable x soit le nombre des résultats favorables d'une série de n épreuves répétées, les probabilités p et q de

l'événement attendu et de l'événement contraire restant les mêmes dans les n épreuves, x est variable d'ordre $n+1$, et la probabilité pour que sa valeur soit k s'écrit $C_n^k p^k q^{n-k}$. La fonction caractéristique

$$\varphi(\theta) = \sum_{k=0}^{k=n} C_n^k p^k q^{n-k} e^{k\theta}$$

s'écrit immédiatement $(q+pe^{\theta})^n$, et son développement en série de puissances de θ prend la forme

$$\left[1+p\left(\frac{\theta}{1}+\frac{\theta^2}{2!}+\dots\right)\right]^n = 1+np\theta+[np+n(n-1)p^2]\frac{\theta^2}{2}+\dots,$$

d'où il suit que

$$\mu_1 = \xi = np,$$
$$m^2 = \mu_2 - \mu_1^2 = np - np^2 = npq,$$

l'écart quadratique moyen relatif à x a donc rigoureusement, et dans tous les cas, la valeur $\sqrt{npq}$ que l'on sait, d'après la théorie élémentaire, être sa valeur asymptotique.

3° Si la fonction $F(x)$ est à *une simple infinité d'escaliers*, la suite $p_0, p_1, p_2, \dots, p_n, \dots$ des probabilités formant une série positive convergente dont la somme est l'unité, la fonction caractéristique

$$\varphi(\theta) = p_0 e^{\theta x_0} + \dots + p_n e^{\theta x_n} + \dots$$

se présente comme une série de fonctions exponentielles. Examinons le problème suivant, traité dans le fascicule I (p. 58) : un très grand nombre de segments de droite égaux consécutifs étant donnés, des points sont placés au hasard sur ces segments de manière qu'il y en ait en moyenne ν sur chaque segment. On démontre que la probabilité pour qu'il y en ait précisément n sur un segment donné est à la limite $p_n = \frac{\nu^n}{n!}e^{-\nu}$.

Si x désigne le nombre des points tombant sur un segment donné, la fonction $F(x)$ correspondante admet les valeurs de discontinuité 0, 1, 2, ..., n, ..., les accroissements $p_0, p_1, \dots, p_n, \dots$ étant les termes de la série

$$e^{-\nu}\cdot 1 + e^{-\nu}\frac{\nu}{1} + \dots + e^{-\nu}\frac{\nu^n}{n!} + \dots = 1.$$

La fonction caractéristique s'écrit

$$e^{-\nu}\left[1+\frac{\nu}{1}e^{\theta}+\frac{\nu^2}{2!}e^{2\theta}+\cdots+\frac{\nu^n}{n!}e^{n\theta}+\cdots\right]$$

où le crochet a pour somme $e^{\nu e^{\theta}}$, donc

$$\varphi(\theta)=e^{\nu(e^{\theta}-1)}$$

Le développement de $\varphi(\theta)$ en série de puissances de θ est

$$1+\nu\frac{\theta}{1}+(\nu+\nu^2)\frac{\theta^2}{2!}+\cdots$$

et l'écart quadratique moyen

$$m=\sqrt{\mu_2-\mu_1^2}$$

a pour valeur $\sqrt{\nu}$, résultat obtenu d'une autre manière dans l'étude du fascicule I signalée plus haut.

4° Dans le cas général, si nous considérons la série

$$1+\frac{\mu_1}{1}\theta+\frac{\mu_2}{2!}\theta^2+\cdots+\frac{\mu_n}{n!}\theta^n+\cdots$$

son rayon de convergence est l'inverse de la « plus grande limite » de $\sqrt[n]{\frac{\mu_n}{n!}}$; si les moments sont tous finis et si l'expression $\sqrt[n]{\frac{\mu_n}{n!}}$ tend vers zéro, la fonction caractéristique est une fonction analytique définie dans tout le plan.

Mais elle est définie pour θ imaginaire pur sans que les moments soient finis, par exemple, dans le cas de la loi, proposée par Cauchy, définie par la fonction des probabilités élémentaires

$$f(x)=\frac{1}{\pi}\,\frac{k}{x^2+k^2},$$

toutes les intégrales

$$\int_{-\infty}^{+\infty}|x|^n f(x)\,dx$$

sont infinies à partir de $n=2$. Cependant, $\varphi(\theta)$ est défini pour θ imaginaire pur : pour $\theta=it$, nous avons, en effet,

$$\varphi(\theta)=\Phi(t)=\frac{k}{\pi}\int_{-\infty}^{+\infty}\frac{e^{ixt}}{x^2+k^2}dx;$$

or, si l'on intègre, pour t positif, la fonction $\frac{e^{itz}}{z^2+k^2}$ le long d'un contour formé, dans le plan de Cauchy, par l'axe réel et un demi-cercle de rayon infini tracé dans le demi-plan supérieur, le résidu relatif au seul pôle existant $z = ki$ est $\frac{e^{-kt}}{2ki}$, l'intégrale obtenue a pour valeur $\frac{\pi}{k}e^{-kt}$; par ailleurs cette intégrale se réduit à sa portion relative à l'axe réel, car si l'on forme le produit $\frac{z}{z^2+k^2}e^{itz}$, pour

$$z = R(\cos\alpha + i\sin\alpha),$$

le module de $\frac{z}{z^2+k^2}$ est nul pour R infini, et celui du terme

$$e^{itz} = e^{-Rt\sin\alpha + iRt\cos\alpha}$$

est nul aussi, puisque t est positif et α compris entre 0 et π.

Il résulte de ce calcul que la fonction caractéristique au sens de M. Paul Lévy a pour expression

$$\Phi(t) = \varphi(t) = e^{-kt},$$

comme la loi de Cauchy est *symétrique*, la fonction caractéristique est évidemment *paire*, donc son expression générale *pour t réel quelconque*, est

$$\Phi(t) = e^{-k|t|}.$$

D'une manière générale, la fonction $\Phi(t)$ est toujours définie pour t réel quelconque, puisque c'est l'intégrale de Stieltjes donnant la valeur moyenne de e^{itx}, fonction continue de x et de module 1, nous voyons là un des avantages de la notion de fonction caractéristique au sens de M. Paul Lévy.

13. La fonction caractéristique détermine la loi de probabilité. — La connaissance de $\Phi(t)$ pour toutes les valeurs réelles de t est équivalente à celle de tous les moments, dans le cas où $\Phi(t)$ est développable en série entière sous la forme

$$\Phi(t) = \mu_0 + it\mu_1 - \frac{t^2}{2!}\mu_2 - i\frac{t^3}{3!}\mu_3 + \cdots \quad (3)$$

Mais on peut aller beaucoup plus loin, et démontrer que, même si le développement (3) n'est pas possible, *la fonction* $\Phi(t)$ *détermine*

complètement la fonction $F(x)$; cette propriété, dans le cas d'une loi absolument continue, se réduit à la *loi de réciprocité* de Fourier, d'après laquelle, si

$$\Phi(t) = \int_{-\infty}^{+\infty} f(x)\, e^{itx}\, dx$$

on a

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi(t)\, e^{-itx}\, dt$$

voici, d'après M. Paul Lévy, sa démonstration dans le cas général.

Le problème consiste à déterminer $F(x)$ par la relation

$$\Phi(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x)$$

or, si l'on considère l'intégrale

$$I_C = \int_{-C}^{C} \Phi(t)\, dt \int_0^X e^{-it\xi}\, d\xi$$

on peut l'exprimer en remplaçant $\Phi(t)$ par sa valeur sous la forme

$$\int_{-C}^{+C} dt \left[\int_0^X e^{-it\xi}\, d\xi\right] \left[\int_{-\infty}^{+\infty} e^{itx}\, dF(x)\right] = \int_{-C}^{C} H(t)\, dt$$

avec

$$H(t) = \int_0^X d\xi \int_{-\infty}^{+\infty} e^{it(x-\xi)}\, dF(x)$$

Cette intégrale double s'écrit, en posant $x - \xi = y$,

$$H(t) = \int_0^X d\xi \int_{-\infty}^{+\infty} e^{ity}\, d_y F(\xi + y),$$

ou, en intervertissant les intégrations,

$$H(t) = \int_{-\infty}^{+\infty} e^{ity}\, dy \int_0^X d_\xi F(\xi + y) = \int_{-\infty}^{+\infty} [F(y + X) - F(y)]\, e^{ity}\, dy.$$

L'intégrale

$$K = \int_\alpha^\beta [F(y + X) - F(y)]\, e^{ity}\, dy$$

a son module moindre que celui de l'intégrale

$$\int_{\alpha}^{\beta} |F(y+\lambda) - F(y)|\, dy$$

qui représente une aire inférieure ou égale à $|\lambda|$, puisque la courbe $u = F(y)$ toute entière, subissant une translation d'amplitude λ parallèlement à l'axe des abscisses, balaye une aire $|\lambda|$. Il s'ensuit que l'intégrale K, quand α et β tendent respectivement vers $+\infty$ et $-\infty$, tend vers sa limite $H(t)$ d'une manière uniforme par rapport à t.

On peut donc, pour le calcul de J_C, intégrer sous le signe somme, et écrire

$$J_C = \int_{-\infty}^{+\infty} [F(y+\lambda) - F(y)] \left| \frac{e^{ity}}{iy} \right|_{-C}^{+C},$$

c'est-à-dire, puisque

$$\left| \frac{e^{ity}}{iy} \right|_{-C}^{+C} = 2\,\frac{\sin Cy}{y},$$

$$J_C = \int_{-\infty}^{+\infty} 2\,\frac{\sin Cy}{y} [F(y+\lambda) - F(y)]\, dy$$

Or, d'après la théorie des *intégrales de Dirichlet*, la limite, pour C infini, de l'intégrale

$$\int_{-\infty}^{+\infty} \frac{\sin Cy}{y}\, \omega(y)\, dy$$

est $2\pi\omega(0)$, si la fonction $\omega(y)$ est continue pour $\omega = 0$, et

$$\pi[\omega(+0) + \omega(-0)]$$

si $\omega(y)$ admet deux valeurs limites distinctes, l'une à droite, l'autre à gauche de $y = 0$. Dans ces conditions, J_C a pour limite

$$J_\infty = 2\pi[F(\lambda) - F(0)],$$

en convenant, toutes les fois que la distribution de masse associée à $F(x)$ présente une masse finie concentrée en un point, d'adopter, pour valeur de $F(x)$, la probabilité pour que la variable soit comprise entre $-\infty$ et x, augmentée de la moitié de la probabilité pour qu'elle soit précisément égale à x

Par ailleurs, le calcul direct de J_C s'effectue aisément; on a

$$J_C = \int_{-C}^{+C} \Phi(t)\, dt\, \frac{1 - e^{-it\lambda}}{it},$$

et, lorsque C tend vers l'infini, la limite de J_C est la *valeur principale*, au sens de Cauchy, de l'intégrale

$$\int_{-\infty}^{+\infty} \frac{1-e^{-itX}}{it}\Phi(t)\,dt$$

La comparaison des deux résultats montre que

$$(4)\qquad F(X)-F(0)=\frac{1}{2\pi}\,vp\int_{-\infty}^{-\infty}\frac{1-e^{-itX}}{it}\Phi(t)\,dt,$$

le symbole vp désignant la « valeur principale » de l'intégrale au sens de Cauchy. Nous voyons que la fonction des probabilités totales est ainsi déterminée, à cela près que F(o) reste inconnu, mais s'obtient immédiatement par la condition $F(-\infty)=0$.

S'il y a une fonction des probabilités élémentaires $f(x)$, on obtient en dérivant la formule (4) par rapport à X

$$(5)\qquad f(X)=\frac{1}{2\pi}\,vp\int_{-\infty}^{-\infty} e^{-itX}\Phi(t)\,dt$$

c'est la formule de Fourier.

14. Exemples. — Afin d'étudier le mécanisme des calculs auxquels conduisent les formules (4) et (5), examinons sommairement deux exemples :

1° Si

$$\Phi(t)=p_1 e^{itx_1}+p_2 e^{itx_2}+\ldots+p_n e^{itx_n}$$

l'intégrale du second membre de (4) porte sur une somme d'expressions de la forme

$$p_k\frac{1-\cos tX+i\sin tX}{it}(\cos t x_k+i\sin t x_k),$$

et se compose ainsi d'une somme S de termes qui se réduisent, pour chaque valeur de k, au produit de p_k par

$$u_k=\int_{-\infty}^{+\infty}\frac{\sin x_k t-\sin(X-x_k)t}{t}\,dt$$

Or, nous savons que l'intégrale

$$\int_{-\infty}^{+\infty} \frac{\sin \lambda t}{t}\, dt$$

est nulle pour $\lambda = 0$, égale à π pour λ positif, à $-\pi$ pour λ négatif, la valeur de u_k dépend donc de la place des nombres x_k et $X - x_k$ par rapport à zéro.

Supposons donc que X passe d'une valeur comprise entre x_{k-1} et x_k à la valeur x_k; le seul terme de la somme S qui change de valeur est le terme

$$p_k \int_{-\infty}^{+\infty} \frac{\sin(X - x_k)t}{t}\, dt,$$

qui passe de la valeur $-\pi p_k$ à la valeur 0. Si X passe ensuite de la valeur x_k à une valeur comprise entre x_k et x_{k+1}, le seul terme qui change de valeur est encore le terme ci-dessus, qui passe de la valeur 0 à la valeur πp_k.

Donc, dans les conditions que nous venons d'envisager, $F(X)$ s'accroît de p_k au total, par deux accroissements égaux chacun à $\frac{p_k}{2}$, il en résulte bien que $F(X)$ est la somme des quantités $p_1, p_2, \ldots$ correspondant à des abscisses $x_1, x_2, \ldots$ inférieures à X, augmentée éventuellement de $\frac{1}{2} p_k$ si $X = x_k$.

2° Appliquons la formule (5) à la fonction caractéristique

$$\Phi(t) = e^{-k|t|}$$

de la loi de Cauchy, nous obtenons

$$f(X) = \frac{1}{2\pi}\int_{-\infty}^{0} e^{(k-iX)t}\, dt + \frac{1}{2\pi}\int_{0}^{+\infty} e^{-(k+iX)t}\, dt$$

ou

$$f(X) = \frac{1}{\pi}\int_{0}^{+\infty} e^{-kt}\cos Xt\, dt,$$

un calcul élémentaire donne bien

$$f(X) = \frac{1}{\pi}\,\frac{k}{k^2 + X^2}.$$

13. Fonction caractéristique d'une somme de plusieurs variables éventuelles — Le problème de la recherche de la loi de probabilité suivie par

$$x = x_1 + x_2 + \dots + x_n$$

connaissant les lois de probabilité, *supposées indépendantes*, de x_1, x_2, ..., x_n, est fondamental, et sa solution grandement facilitee par l'emploi des fonctions caractéristiques Quel que soit α, réel ou imaginaire, l'exponentielle

$$e^{\alpha x} = e^{\alpha(x_1 + x_2 + \dots + x_n)}$$

est le produit des exponentielles $e^{\alpha x_1}$, $e^{\alpha x_2}$, ..., $e^{\alpha x_n}$ Donc l'intégrale multiple d'ordre n

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{\alpha x_1} e^{\alpha x_2} \dots e^{\alpha x_n} \, dF(x_1) \, dF(x_2) \dots dF(x_n)$$

est le produit des intégrales simples $\int_{-\infty}^{+\infty} e^{\alpha x} dF(x_h)$ Il s'ensuit que *la fonction caractéristique de la loi de probabilité relative à x (loi résultante) est le produit des fonctions caractéristiques des lois relatives à x_1, x_2, ..., x_n (lois composantes)*

Si les développements en série du type (1) sont valables, nous avons en nous limitant à deux variables composantes x' et x'', et posant $x = x' + x''$,

$$\mu_0 + \frac{\mu_1}{1}\theta + \frac{\mu_2}{2!}\theta^2 + \dots$$
$$= \left[\mu'_0 + \frac{\mu'_1}{1}\theta + \frac{\mu'_2}{2!}\theta^2 + \dots\right]\left[\mu''_0 + \frac{\mu''_1}{1}\theta + \frac{\mu''_2}{2!}\theta^2 + \dots\right]$$

Il en résulte en égalant les coefficients des puissances successives de θ

$$\mu_0 = \mu'_0 \mu''_0, \qquad \text{ou du reste} \quad \mu_0 = \mu'_0 = \mu''_0 = 1,$$
$$\mu_1 = \mu'_1 + \mu''_1,$$
$$\mu_2 = \mu'_2 + 2\mu'_1 \mu''_1 + \mu''_2 = (\mu'_1 + \mu''_1)^{(2)},$$

la puissance symbolique signifiant que μ'^2_1 et μ''^2_1 sont remplacés par μ'_2 et μ''_2 ; et d'une manière générale

$$\mu_h = (\mu'_1 + \mu''_1)^{(h)}$$

Si les lois composantes sont *toutes deux symétriqu*

d'ordre impair sont tous nuls, et l'on a, en ce qui concerne les moments

d'ordres 3 et 4,

$$\mu_2 = \mu'_2 + \mu''_2,$$
$$\mu_4 = \mu'_4 + 6\mu'_2\mu''_2 + \mu''_4.$$

Le résultat relatif aux moments d'ordre 2 est très important, et vient à l'appui de ce que nous avons dit au § 8 sur le rôle capital de l'écart quadratique moyen. Il s'étend à un nombre quelconque de lois composantes symétriques, si $x_1, x_2, \ldots, x_n$ obéissent à de telles lois, le carré moyen de $x = x_1 + x_2 + \ldots + x_n$ est la somme des carrés moyens de $x_1, x_2, \ldots, x_n$.

CHAPITRE III.

LE PROBLÈME GÉNÉRAL DES ERREURS D'OBSERVATION.

16. **Erreurs accidentelles et erreurs systématiques.** — On mesure une grandeur dont la valeur véritable est Z, la mesure donne le résultat z, la différence $Z - z$ est l'erreur commise, que nous considérons donc comme positive si elle est par défaut, négative si elle est par excès. La mesure répétée n fois, dans les mêmes conditions, donne les résultats $z_1, z_2, \ldots, z_n$, entachés des erreurs $e_1, e_2, \ldots, e_n$. Nous supposons que les n mesures ont été faites dans des conditions en tous points semblables, en particulier que c'est bien *la même grandeur*, parfaitement déterminée, qui a été mesurée, il peut y avoir là l'origine de difficultés sur lesquelles nous ne nous arrêterons pas.

Les erreurs $e_1, e_2, \ldots, e_n$ sont les sommes algébriques de deux sortes d'erreurs : les erreurs *systématiques* et les erreurs *accidentelles*.

Une erreur systématique a une cause déterminée qui fausse toujours dans le même sens, le résultat de la mesure : le mètre servant à mesurer une longueur est trop court, la balance servant à mesurer une masse a ses bras de leviers inégaux, etc. On s'efforce, dans les techniques des diverses catégories de mesures, de déterminer les causes d'erreurs systématiques et d'annihiler leurs effets par des corrections (en thermométrie par exemple).

Les erreurs accidentelles dépendent *du hasard*, chacune d'elles est la somme algébrique d'un grand nombre de petites erreurs dont les causes sont nombreuses et peuvent généralement agir dans l'un ou l'autre sens : par exemple, si la mesure aboutit à une lecture relative à un spot lumineux, les petites variations dans l'atmosphère traversée par les rayons.

La distinction établie ainsi entre les deux catégories d'erreurs est

toujours sujette à critique et à revision, un perfectionnement dans la technique d'une mesure aboutit souvent à la mise en évidence de nouvelles causes d'erreurs systématiques et par conséquent à une épuration des erreurs subsistant après les corrections et qualifiées d'accidentelles.

De plus, il peut se glisser dans une série de mesures des erreurs isolées, souvent grossières, comme l'emploi par inadvertance, dans une pesée, d'un poids de cent grammes pour un de cinquante. On ne saurait tenir compte, dans une théorie des erreurs, des cas rédhibitoires de cette nature, et d'ailleurs l'emploi des méthodes que nous étudierons permet de discerner et d'éliminer les observations ainsi entachées d'erreurs parasites notables.

Si les corrections sont parfaites, hypothèse évidemment idéale, l'erreur commise est purement accidentelle, on doit la considérer comme une *variable éventuelle*, et l'un des premiers objets de l'étude des erreurs d'observation dans une catégorie de mesures n'est autre que la loi de probabilité de cette variable.

17. La loi de Gauss. — Cette étude se trouve dominée et conditionnée par une loi de probabilité classique, qui a une importance théorique et pratique fondamentale, la *loi de Gauss*, caractérisée par la fonction des probabilités élémentaires

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2},$$

relative à l'erreur $x = Z - z$.

Cette loi est identique à la loi normale des écarts dans la théorie des épreuves répétées. Elle a été l'objet de diverses *justifications théoriques* que nous étudierons plus loin, et ses conséquences sont généralement considérées comme confirmées d'une manière satisfaisante par l'expérience. Cependant les justifications s'appuient sur diverses hypothèses qui peuvent être soumises à la critique, et la question des vérifications expérimentales mérite peut-être d'être reprise par les méthodes les plus modernes de la *statistique mathématique*. Il demeure certain que, par sa simplicité et ses propriétés remarquables, la loi de Gauss doit rester universelle tout au moins en première approximation.

Nous lui consacrerons toute une partie de ce fascicule, pour étudier

les diverses justifications qui en ont été données ainsi que ses principales applications à la « combinaison des observations ». Nous reprendrons ensuite les problèmes pratiques ainsi traités, et nous verrons, d'après les travaux de Gauss lui-même, que les méthodes auxquelles conduit la loi ci-dessus peuvent se justifier, d'un point de vue différent, sans admettre aucune forme analytique précise pour les lois de probabilité des erreurs commises dans les observations mises en cause. Le but du présent Chapitre est d'énoncer auparavant d'une manière sommaire les principaux problèmes pratiques que nous devons étudier.

18. **Combinaison d'observations directes.** — Le premier problème posé par la théorie des erreurs d'observation est en fait le suivant :

On a effectué n déterminations expérimentales d'une même grandeur, et obtenu ainsi les résultats

$$z_1, \quad z_2, \quad \ldots, \quad z_n;$$

quelle valeur convient-il d'adopter pour la grandeur mesurée, et de combien peut-on craindre que la valeur ainsi adoptée s'écarte de la valeur véritable ?

Si l'on admet que les erreurs commises sont des erreurs purement accidentelles, et que les mesures ont été faites dans des conditions entièrement semblables, on est fondé à conclure que la loi de probabilité des erreurs est absolument continue, et que la fonction des probabilités élémentaires dépend seulement de la valeur véritable Z et de la valeur mesurée z; soit

$$f(z, Z)$$

cette fonction.

Supposant connue la loi de probabilité ainsi représentée, on peut considérer le problème posé comme un problème de *probabilité des causes*. Désignons par $\varphi(Z)\,dZ$ la probabilité *a priori* pour que la grandeur mesurée ait une valeur véritable comprise entre Z et $Z + dZ$; la probabilité *a posteriori* sera, d'après le § 5,

$$P\,dZ = \frac{\varphi(Z) f(Z, z_1) f(Z, z_2) \ldots f(Z, z_n)\,dZ}{\int_{-\infty}^{+\infty} \varphi(Z) f(Z, z_1) f(Z, z_2) \ldots f(Z, z_n)\,dZ}.$$

Le choix de la valeur qu'il convient d'adopter pour Z peut se faire, dès lors, selon divers points de vue; on peut, notamment, choisir la *valeur la plus probable*, ou encore la *valeur moyenne*. En tout cas, dans l'ignorance où l'on se trouve de la nature de la fonction $\varphi(Z)$, on ne peut que la supposer constante, selon *l'hypothèse de Bayes;* il en résulte une impression de doute sur la solidité mathématique de la méthode ainsi adoptée. Nous verrons qu'on peut s'affranchir de cette hypothèse et même de la théorie des probabilités des causes si l'on admet que les erreurs sont purement accidentelles et très petites, en adoptant la valeur de Z sur laquelle l'erreur quadratique moyenne est minimum (Chap. IX).

19. Combinaison d'observations indirectes. — Le problème de la combinaison d'observations indirectes généralise le précédent; il consiste à déterminer les valeurs les plus plausibles de certaines grandeurs inconnues X, Y, Z, ..., d'après les mesures directes d'autres grandeurs qui sont des fonctions de ces inconnues.

Soient donc k grandeurs inconnues X, Y, Z, ..., et soient L_1, L_2, ..., L_n les résultats des mesures directes des quantités

$$
\begin{aligned}
U_1 &= f_1(X, Y, Z, \ldots),\\
U_2 &= f_2(X, Y, Z, \ldots),\\
&\ldots\\
U_n &= f_n(X, Y, Z, \ldots).
\end{aligned}
$$

Le plus souvent, les fonctions ainsi mises en jeu sont les formes prises par une même fonction des inconnues X, Y, Z, ..., et de certains paramètres, en donnant diverses valeurs à ces derniers paramètres.

Supposant tout d'abord que les inconnues X, Y, Z, ..., sont *indépendantes*, c'est-à-dire qu'aucune relation tenant à la nature même du phénomène étudié n'existe entre elles, nous nous proposons de chercher les valeurs qu'il convient d'adopter pour les inconnues, et les erreurs que l'on peut craindre sur les valeurs ainsi adoptées.

Cet énoncé suppose naturellement n supérieur à k. Les observations faites comportent des erreurs, de sorte que les équations

$$
\begin{aligned}
f_1(X, Y, Z, \ldots) &= L_1,\\
\ldots\ldots &\ldots\\
f_n(X, Y, Z, \ldots) &= L_n.
\end{aligned}
$$

ne sont pas compatibles en X, Y, Z, ... Le problème de la recherche de la solution approchée la plus avantageuse du système qu'elles forment est considéré par Gauss comme le plus important parmi ceux que présente l'application des mathématiques à la philosophie naturelle

On peut interpréter d'une manière assez arbitraire l'expression de « solution la plus avantageuse » ; une interprétation bien naturelle consiste à adopter les *valeurs les plus probables* des inconnues.

Cherchons donc la probabilité pour que, vu les résultats des observations, les inconnues aient des valeurs respectivement comprises entre X et $X + dX$, Y et $Y + dY$, Z et $Z + dZ$, ... nous avons affaire à un problème de probabilité des causes.

Soit

$$\varphi(X, Y, Z, \ldots) dX\, dY\, dZ \ldots$$

la probabilité élémentaire *a priori* relative aux valeurs des inconnues délimitées ainsi qu'il est dit ci-dessus La cause d'après laquelle il en est ainsi étant supposée mise en jeu, la probabilité pour que les observations aient donné les résultats $L_1, L_2, \ldots, L_n$, c'est-à-dire les erreurs

$$f_1(X, Y, Z \ldots) - L_1, \quad \ldots, \quad f_n(X, Y, Z \ldots) - L_n$$

est proportionnelle au produit des fonctions des probabilités élémentaires

$$\theta_1(U_1, L_1), \quad \theta_2(U_2, L_2), \quad \ldots \quad \theta_n(U_n, L_n)$$

caractérisant les lois respectives d'erreurs des observations

D'après le § 5, la probabilité *a posteriori* pour que les inconnues aient des valeurs véritables comprises entre X et $X + dX$, Y et $Y + dY$, ... est donc

$$\frac{\varphi\,\theta_1\,\theta_2 \ldots \theta_n\, dX\, dY\, dZ \ldots}{\underbrace{\int \ldots \int}_{k} \varphi\,\theta_1\,\theta_2 \ldots \theta_n\, dX\, dY\, dZ \ldots},$$

l'intégrale du dénominateur étant etendue à tout le domaine de l'espace (X, Y, Z, ...) représentant l'ensemble des systèmes de valeurs des inconnnues considérés comme possibles *a priori.*

Cette probabilité élémentaire est proportionnelle au produit

$$\varphi\,\theta_1\,\theta_2 \ldots \theta_n,$$

donc le système des valeurs les plus probables des inconnues sera celui qui rendra ce produit maximum.

L'expression de la probabilité *a posteriori* calculée plus haut permettra aussi, en principe, d'envisager le problème en considérant comme valeurs les plus plausibles les *valeurs probables* des inconnues.

Dans l'un et l'autre cas, la solution exige la connaissance des fonctions $\theta_1, \theta_2, \ldots, \theta_n$ caractérisant les lois de probabilité des erreurs commises, et d'autre part l'adoption d'une hypothèse sur la fonction $\varphi(X, Y, Z, \ldots)$, en fait, faute de tout renseignement, on adopte encore l'hypothèse de Bayes. Nous verrons, pour ce problème ainsi que pour le précédent, comment on peut traiter la question en s'affranchissant de ces diverses conditions, au moyen du minimum des erreurs quadratiques moyennes (Chap. X).

20. Compensation d'observations conditionnelles. — Supposons maintenant que les inconnues $X, Y, Z, \ldots$ dont la détermination résulte des mesures faites sur les quantités $U_1, U_2, \ldots, U_n$ ne sont plus indépendantes, mais sont liées par h relations

$$\begin{aligned} \varphi_1(X, Y, Z, \ldots) &= 0 \\ \varphi_2(X, Y, Z, \ldots) &= 0 \\ &\cdots \\ \varphi_h(X, Y, Z, \ldots) &= 0, \end{aligned}$$

tenant à la nature même du problème proposé, nous supposons naturellement $k > h$ et $n > k - h$. Le problème de la détermination des valeurs les plus plausibles des inconnues, parmi toutes les solutions exactes du système des équations ci-dessus, peut aisément se ramener au précédent, car on peut toujours admettre que l'on résout les équations $\varphi_1 = 0, \varphi_2 = 0, \ldots, \varphi_h = 0$ par rapport à h inconnues, que l'on exprime en fonction de $q = k - h$ inconnues principales indépendantes. En désignant par $X_1, X_2, \ldots, X_q$ ces inconnues, on peut reprendre l'étude du paragraphe précédent, et mettre sous la forme

$$\frac{\psi\theta_1\theta_2\ldots\theta_n\, dX_1\, dX_2\ldots dX_q}{\underbrace{\int\cdots\int}_{q}\psi\theta_1\theta_2\ldots\theta_n\, dX_1\, dX_2\ldots dX_q}$$

la probabilité élémentaire *a posteriori* relative aux valeurs X_1, X_2, ..., X_q. Connaissant les lois des erreurs et admettant encore l'hypothèse $\psi = \text{const}$, on peut déterminer les valeurs les plus probables des inconnues ou leurs valeurs probables.

Mais le problème des observations conditionnelles peut être envisagé d'une manière différente. En éliminant les inconnues X, Y, Z, ..., entre les relations $U_1 = f_1$, ..., $U_n = f_n$, et les relations $\varphi_1 = 0$, $\varphi_2 = 0$, ..., $\varphi_h = 0$, on obtient, entre les quantités U_1, U_2, ..., U_n, des relations, au nombre de $p = n + h - k = n - q$, de la forme

$$\begin{aligned} g_1(U_1, U_2, \ldots, U_n) &= 0, \\ g_2(U_1, U_2, \ldots, U_n) &= 0, \\ &\ldots \\ g_p(U_1, U_2, \ldots, U_n) &= 0, \end{aligned}$$

et que l'on appelle les *équations de condition*. Le problème consistant à trouver les valeurs les plus plausibles des quantités mesurées, ces valeurs devant vérifier les équations de condition ci-dessus, est le problème de la *compensation* des observations conditionnelles. Si les résultats de mesure sont L_1, L_2, ..., L_n, les valeurs adoptées, dites *valeurs compensées*, sont $L_1 + \lambda_1$, $L_2 + \lambda_2$, ..., $L_n + \lambda_n$, et les quantités λ_1, λ_2, ..., λ_n sont les *corrections*.

Ce deuxième point de vue s'applique même si les inconnues X, Y, Z, ... sont indépendantes, le nombre des équations de condition est seulement égal alors à $n - k$. En définitive, il y a lieu de distinguer *deux méthodes distinctes*, et non pas deux problèmes distincts. Nous étudierons la méthode de compensation, très employée notamment en *géodésie*, d'abord en admettant la loi de Gauss et l'hypothèse de Bayes (Chap. VIII, § 3), puis en nous affranchissant de toute hypothèse sur les fonctions ψ, θ_1, θ_2, ..., θ_n et nous plaçant au point de vue du minimum des erreurs quadratiques moyennes (Chap. XII).

DEUXIÈME PARTIE.

LA LOI DE GAUSS.

CHAPITRE IV.

PROPRIÉTÉS GÉNÉRALES DE LA LOI DE GAUSS [1]

21. **La fonction Θ; la courbe en cloche** — La loi de Gauss, énoncée au § 17, consiste en ce que la probabilité pour que l'erreur commise dans une observation soit comprise entre x et $x+dx$ a pour expression

$$\frac{k}{\sqrt{\pi}} e^{-k^2 x^2} dx,$$

nous devons, avant d'aborder l'étude des justifications théoriques qui en ont été données par Gauss et depuis Gauss, passer en revue ses propriétés les plus importantes.

La fonction des probabilités totales correspondante est

$$F(x) = \frac{k}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-k^2 x^2} dx,$$

elle est moins employée, dans les nombreux travaux traitant de la question, que la fonction exprimant la probabilité d'une erreur comprise entre $-\alpha$ et $+\alpha$.

[1] Nous adoptons ici la dénomination la plus couramment employée en France pour la loi classique des erreurs, que les auteurs américains appellent souvent « deuxième loi de Laplace » et que les mathématiciens anglais font remonter à DE MOIVRE L'intégrale du haut de la page suivante, en tout état de cause, doit être appelée « intégrale de Laplace »

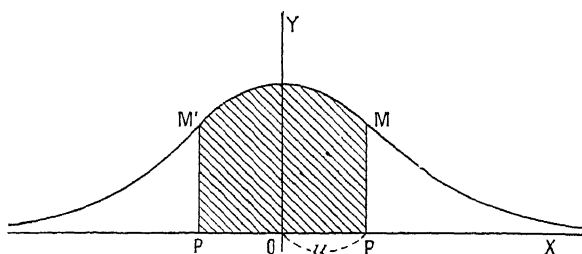
En posant $h\varepsilon = u$, cette probabilité

$$\frac{h}{\sqrt{\pi}}\int_{-\varepsilon}^{+\varepsilon} e^{-h^2x^2}\,dx$$

est représentée par l'*intégrale de Laplace*

$$\Theta(u) = \frac{2}{\sqrt{\pi}}\int_0^u e^{-X^2}\,dX$$

qui mesure l'aire comprise entre les droites $X = u$, $X = -u$, l'axe OX, et la *courbe en cloche* d'équation $Y = \frac{1}{\sqrt{\pi}}e^{-X^2}$ (*fig.* 3).

Fig. 3

Les *tables numériques* de la fonction Θ constituent un instrument de travail fondamental; le lecteur en trouvera une en fin du présent volume, donnant $\Theta(u)$ avec cinq décimales pour u variant de millième en millième entre 0 et 1,600, puis de centième en centième entre 1,60 et 3,18.

D'après la formule $h\varepsilon = u$, l'erreur qu'on a une probabilité donnée de ne pas dépasser en valeur absolue, c'est-à-dire qui correspond à une valeur donnée de u, est *inversement proportionnelle au paramètre* h. Nous appellerons h le *paramètre de précision* de la série de mesures considérée.

22. Moments et fonction caractéristique de la loi de Gauss. — Rappelons d'abord le calcul classique de l'intégrale

$$J = \int_{-\infty}^{+\infty} e^{-ax^2}\,dx,$$

pour a positif donné.

L'intégrale double

$$H = \int\int e^{-a(x^2+y^2)}\,dx\,dy$$

étendue à tout le plan, est égale à J^2, et sa valeur, calculée en coordonnées polaires, est

$$\int_0^{2\pi} d\theta \int_0^{+\infty} e^{-a\rho^2}\rho\,d\rho = 2\pi\left| -\frac{1}{2a}e^{-a\rho^2}\right|_0^{\infty},$$

c'est-à-dire

$$H = \frac{\pi}{a}.$$

Par conséquent

$$\int_{-\infty}^{+\infty} e^{-ax^2}\,dx = \frac{\sqrt{\pi}}{\sqrt{a}}. \tag{1}$$

Les intégrales

$$\int_{-\infty}^{+\infty} e^{-a(x-h)^2}\,dx \quad \text{et} \quad \int_{-\infty}^{+\infty} e^{-a(x+ih)^2}\,dx$$

h étant réel, ont aussi pour valeur $\frac{\sqrt{\pi}}{\sqrt{a}}$. En effet, la première se ramène à J par le changement de variable $x = h + u$. Quant à la deuxième, c'est, dans le plan de la variable complexe, l'intégrale de la fonction entière $Z = e^{-az^2}$ le long de la droite (D) d'équation $y = h$, parcourue en faisant varier x de $-\infty$ à $+\infty$; le module de Z étant nul aux points à l'infini sur Ox, l'intégrale a la même valeur le long de (D) et le long de Ox.

Moments de la loi de Gauss. — Faisant $a = k^2$ dans la formule (1), nous avons $\mu_0 = 1$, résultat classique et du reste conforme à la nature même de la question.

Dérivant p fois l'identité (1) par rapport à a, nous avons ensuite

$$(-1)^p \int_{-\infty}^{+\infty} x^{2p} e^{-ax^2}\,dx = \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2p-1}{2}\right)\frac{\sqrt{\pi}}{a^{p+\frac{1}{2}}};$$

il en résulte en prenant toujours $a = k^2$

$$\mu_{2p} = \frac{k}{\sqrt{\pi}}\int_{-\infty}^{+\infty} x^{2p} e^{-k^2x^2}\,dx = \frac{1 \cdot 3 \cdot 5 \ldots 2p-1}{2^p k^{2p}};$$

en particulier,

$$\mu_2 = \frac{1}{2k^2}, \qquad \mu_4 = \frac{3}{4k^4}.$$

Les moments d'ordre impair sont tous nuls, puisque la loi de probabilité est symétrique; mais le calcul des quantités

$$\lambda_{2p+1} = \frac{k}{\sqrt{\pi}} \int_{-\infty}^{+\infty} |x|^{2p+1} e^{-k^2x^2}\,dx$$

a son importance.

Partant de l'identité

$$2\int_0^{+\infty} x e^{-ax^2}\,dx = \frac{1}{a},$$

qui nous donne

$$\lambda_1 = \frac{2k}{\sqrt{\pi}} \int_0^{+\infty} x e^{-k^2x^2}\,dx = \frac{1}{k\sqrt{\pi}},$$

nous avons, en dérivant p fois par rapport à a,

$$(-1)^p\,2\int_0^{+\infty} x^{2p+1} e^{-ax^2}\,dx = (-1)(-2)\ldots(-p)\frac{1}{a^{p+1}},$$

et il en résulte

$$\lambda_{2p+1} = \frac{p!}{k^{2p+1}\sqrt{\pi}}.$$

Fonction caractéristique. — Le calcul direct de

$$\varphi(\theta) = \frac{k}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{\theta x} e^{-k^2x^2}\,dx$$

se fait sans difficulté, l'exposant

$$\theta x - k^2 x^2$$

s'écrit en effet

$$\frac{\theta^2}{4k^2} - \left(kx - \frac{\theta}{2k}\right)^2,$$

et nous avons, par le changement de variable $x = \frac{\theta}{2k^2} + u$,

$$\varphi(\theta) = \frac{k}{\sqrt{\pi}} e^{\frac{\theta^2}{4k^2}} \int_{-\infty}^{+\infty} e^{-k^2u^2}\,du,$$

donc

$$\varphi(\theta) = e^{\frac{\theta^2}{4k^2}}.$$

Le développement de $\varphi(\theta)$ sous la forme

$$1+\frac{1}{4k^2}\theta^2+\ldots+\frac{1}{p!}\,\frac{1}{2^{2p}k^{2p}}\theta^{2p}+\ldots$$

permet de retrouver directement les expressions des moments

$$\mu_{2p}=\frac{2p!}{2^{2p}\,p!}\,\frac{1}{k^{2p}}=\frac{1\cdot 3\cdot 5\ldots 2p-1}{2^p k^{2p}};$$

quant à la fonction

$$\Phi(t)=\varphi(it),$$

son expression sous la forme

$$1-\frac{\mu_2}{2!}t^2+\frac{\mu_4}{4!}t^4-\ldots$$

donne le résultat

$$\Phi(t)=e^{-\frac{t^2}{4k^2}}.$$

Le calcul direct, au moyen de l'intégrale

$$\Phi(t)=\frac{k}{\sqrt{\pi}}\int_{-\infty}^{+\infty}e^{itx}\,e^{-k^2x^2}\,dx$$

se ramène à celui de

$$\int_{-\infty}^{+\infty}e^{-a(x+b)^2}\,dx=\frac{\sqrt{\pi}}{\sqrt{a}};$$

l'exposant s'écrit en effet

$$itx-k^2x^2=-\frac{t^2}{4k^2}-\left(kx-\frac{it}{2k}\right)^2,$$

et l'on a

$$\Phi(t)=e^{-\frac{t^2}{4k^2}}\,\frac{k}{\sqrt{\pi}}\int_{-\infty}^{+\infty}e^{-k^2\left(x-\frac{it}{2k^2}\right)^2}dx=e^{-\frac{t^2}{4k^2}}.$$

La formule de réciprocité de Fourier reproduit du reste bien

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-ixt}\,\Phi(t)\,dt=\frac{1}{2\pi}e^{-k^2x^2}\int_{-\infty}^{+\infty}e^{-\frac{1}{4k^2}(t+2ik^2x)^2}\,dt,$$

qui se réduit à

$$\frac{k}{\sqrt{\pi}}e^{-k^2x^2}.$$

23. Erreur absolue moyenne, erreur médiane; erreur quadratique moyenne. — La quantité $\lambda_1 = \frac{1}{k\sqrt{\pi}}$, calculée plus haut, est la valeur moyenne du module de l'erreur, on la nomme souvent *l'erreur absolue moyenne;* elle est à l'erreur quadratique moyenne $m = \frac{1}{k\sqrt{2}}$ dans le rapport $\frac{\sqrt{2}}{\sqrt{\pi}}$, et ce résultat est susceptible de vérification expérimentale simple.

On peut caractériser la précision de la catégorie de mesures considérée par la donnée de divers éléments numériques autres que le paramètre de précision k, et notamment par ce que M. Fréchet appelle des *valeurs typiques* de l'erreur. L'erreur absolue moyenne et l'erreur quadratique moyenne sont des erreurs typiques. On utilise aussi comme erreur typique *l'erreur médiane*, c'est-à-dire l'erreur qu'on a une probabilité $\frac{1}{2}$ de ne pas dépasser en valeur absolue, sa valeur est $\frac{u}{k}$, le nombre u correspondant étant défini par la condition

$$\Theta(u) = \frac{1}{2};$$

il a pour valeur approchée 0,47693; les valeurs de u correspondant à l'erreur quadratique moyenne et l'erreur absolue moyenne sont respectivement $\frac{1}{\sqrt{2}} = 0,70710$ environ, et $\frac{1}{\sqrt{\pi}} = 0,56419$ environ. Si l'on désigne par les lettres λ, m, μ les erreurs absolue moyenne, quadratique moyenne et médiane, on a très approximativement

$$k = \frac{0,56419}{\lambda} = \frac{0,70710}{m} = \frac{0,47693}{\mu},$$

$$\frac{m}{\lambda} = 1,25331, \qquad \frac{\mu}{\lambda} = 0,84533, \qquad \frac{\mu}{m} = 0,67449.$$

Les probabilités pour que l'erreur ne dépasse pas, en valeur absolue, les valeurs λ, m, μ sont respectivement

$$\Theta\left(\frac{1}{\sqrt{\pi}}\right) = 0,5750 \text{ environ}, \qquad \Theta\left(\frac{1}{\sqrt{2}}\right) = 0,6826 \text{ environ, et } \frac{1}{2}.$$

Dans les premières pages du présent Chapitre, nous avons conservé le *paramètre de précision* k qui est utilisé dans de nombreux

Ouvrages et Mémoires importants; mais il y a un avantage certain, que la suite de ce Volume mettra en évidence à exprimer les résultats fondamentaux en fonction de *l'erreur quadratique moyenne m*.

La fonction des probabilités élémentaires

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$$

devient ainsi

$$f(x) = \frac{1}{m\sqrt{2\pi}} e^{-\frac{x^2}{2m^2}},$$

la fonction caractéristique $\Phi(t)$ s'écrit

$$\Phi(t) = e^{-\frac{m^2t^2}{2}}$$

et les moments ont pour expressions

$$\mu_2 = m^2, \quad \mu_4 = 3m^4, \quad \ldots, \quad \mu_{2p} = 1.3.5\ldots(2p-1)m^{2p},$$

enfin les valeurs moyennes des puissances impaires de x sont

$$\lambda_1 = m\sqrt{\frac{2}{\pi}}, \quad \ldots, \quad \lambda_{2p+1} = 2.4.6\ldots(2p)\sqrt{\frac{2}{\pi}}\, m^{2p+1}.$$

24. Réduction d'une loi de probabilité. — On dit qu'une loi de probabilité est *réduite* lorsqu'elle est exprimée en fonction de la variable

$$\xi = \frac{x - \mu_1}{m},$$

où μ_1 désigne la *valeur moyenne* de x, et m *l'écart quadratique moyen* $\sqrt{\mu_2 - \mu_1^2}$ (*voir* § 8).

Il est clair que, si l'on désigne par $\overline{F}(\xi)$ la fonction des probabilités totales de la loi réduite, on a

$$\overline{F}(\xi) = F(x);$$

il en résulte donc, s'il existe une fonction des probabilités élémentaires, que

$$\overline{f}(\xi) = \frac{d\overline{F}}{d\xi} = m\frac{dF}{dx} = m f(x)$$

par exemple, dans le cas de la loi de Gauss, on a

$$\bar{f}(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}.$$

Quant à la fonction caractéristique de la loi réduite, que nous désignerons par $\bar{\Phi}(\tau)$, et qui est la valeur moyenne de $e^{i\tau\xi}$, son expression est

$$\bar{\Phi}(\tau) = \int_{-\infty}^{+\infty} e^{i\tau\xi}\, d\bar{F}(\xi) = e^{-\frac{\mu_1}{m} i\tau} \int_{-\infty}^{+\infty} e^{ix\frac{\tau}{m}}\, dF(x),$$

c'est-à-dire

$$\bar{\Phi}(\tau) = e^{-\frac{\mu_1}{m} i\tau}\, \Phi\left(\frac{\tau}{m}\right),$$

on a inversement, en posant $\tau = mt$,

$$\Phi(t) = e^{\mu_1 it}\, \bar{\Phi}(mt).$$

La fonction caractéristique de la loi de Gauss réduite est, dans ces conditions,

$$\bar{\Phi}(\tau) = e^{-\frac{\tau^2}{2}}.$$

Nous aurons l'occasion d'utiliser, au Chapitre VII, le mécanisme que nous venons d'étudier pour la formation des fonctions $\bar{F}$, $\bar{f}$, $\bar{\Phi}$, à partir de F, f, Φ, ou inversement.

25. La loi de Gauss et la théorie des epreuves répétées. — On peut utiliser les notions qui précèdent et les considérations du § 15 sur la composition des lois de probabilité pour déduire de la notion de fonction caractéristique une confirmation de la théorie classique des épreuves répétées.

Si nous considérons la variable éventuelle x, écart entre le nombre de résultats favorables d'une série de n épreuves répétées et la valeur probable np de ce nombre (p et q désignant toujours les probabilités de l'événement attendu et de l'événement contraire), nous avons

$$x = x_1 + x_2 + \ldots + x_n,$$

où x_1 désigne l'écart analogue pour la première épreuve, x_2 pour la deuxième, etc.

Si l'événement attendu se produit, on a $x_1 = q$, et la probabilité est p; si c'est l'événement contraire, $x_1 = -p$ et la probabilité est q.

La fonction caractéristique s'écrit donc

$$\Phi_1(t) = p\, e^{iqt} + q\, e^{-ipt},$$

pour la variable x, le produit $\Phi_1 \Phi_2 \ldots \Phi_n = \Phi_1^n$ donne

$$\Phi(t) = [p\, e^{iqt} + q\, e^{-ipt}]^n,$$

ou, en développant suivant les puissances de t, et tenant compte de la relation $p + q = 1$,

$$\Phi(t) = \left[1 - \frac{pq}{2} t^2 + ipq(p-q)\frac{t^3}{3!} + pq(1-3pq)\frac{t^4}{4!} + \ldots\right]^n.$$

La valeur quadratique moyenne de x est

$$m = \sqrt{\mu_2} = \sqrt{npq},$$

et, si l'on envisage la loi réduite correspondante, c'est-à-dire, puisque $\mu_1 = 0$, la loi de probabilité de

$$\xi = \frac{x}{\sqrt{npq}},$$

la fonction caractéristique de cette loi réduite est

$$\overline{\Phi(\tau)} = \int_{-\infty}^{+\infty} e^{\frac{i\tau x}{\sqrt{npq}}} dF(x) = \Phi\left(\frac{\tau}{\sqrt{npq}}\right)$$

ou, d'après l'expression ci-dessus de $\Phi(t)$,

$$\overline{\Phi(\tau)} = \left[1 - \frac{\tau^2}{2n} + i\frac{p-q}{n\sqrt{npq}}\frac{\tau^3}{3!} + \frac{1-3pq}{n^2pq}\frac{\tau^4}{4!} + \ldots\right]^n$$

Le logarithme de $\overline{\Phi(\tau)}$ admet par conséquent le développement

$$\operatorname{Log} \overline{\Phi(\tau)} = \overline{\Psi(\tau)} = -\frac{\tau^2}{2} + i\frac{p-q}{\sqrt{npq}}\frac{\tau^3}{3!} + \ldots$$

dont tous les termes, à partir du second, contiennent en dénominateur les puissances successives de $\sqrt{n}$. *Lorsque n augmente indéfiniment, $\overline{\Phi(\tau)}$ tend vers $-\frac{\tau^2}{2}$, et $\overline{\Phi(\tau)}$ vers $e^{-\frac{\tau^2}{2}}$, fonction caractéristique de la loi de Gauss réduite.*

On peut aisément étendre cette étude au cas de

$$N = n_1 + n_2 + \ldots + n_k,$$

épreuves dont les n_1 premières sont relatives à un événement de probabilité p_1, ..., les n_k derniers à un evenement de probabilité p_k. Soit

$$x = x_1 + x_2 + \cdots + x_k$$

l'écart entre le nombre de résultats favorables et sa valeur moyenne

$$n_1p_1 + n_2p_2 + \cdots + n_kp_k.$$

La fonction caractéristique relative à x_1 est

$$(p_1 e^{iq_1t} + q_1 e^{-ip_1t})^{n_1} = \left[1 - \frac{p_1q_1}{2}t^2 + i\frac{p_1q_1(p_1-q_1)}{3!}t^3 + \cdots\right]^{n_1}$$

et la fonction caracteristique relative à x est le produit des k expressions analogues.

On obtient ainsi

$$\Phi(t) = 1 - \frac{t^2}{2}\Sigma n_i p_i q_i + \cdots$$

et l'on voit que la valeur quadratique moyenne de x est

$$m = \sqrt{n_1p_1q_1 + \cdots + n_kp_kq_k},$$

de sorte que la loi réduite a pour fonction caractéristique

$$\overline{\Phi(\tau)} = \Phi\left(\frac{\tau}{m}\right) = \Phi\left(\frac{\tau}{\sqrt{\Sigma n_h p_h q_h}}\right),$$

et que l'on a

$$\overline{\Psi(\tau)} = \text{Log}\,\overline{\Phi(\tau)} = \Sigma n_i\,\text{Log}\left[1 - \frac{p_iq_i}{2\Sigma n_h p_h q_h}\tau^2 + i\frac{p_iq_i(p_i-q_i)}{3!(\Sigma n_h p_h q_h)^{\frac{3}{2}}}\tau^3 + \cdots\right]$$

ou

$$-\frac{\tau^2}{2} + i\frac{\Sigma n_i p_i q_i (p_i - q_i)}{(\Sigma n_h p_h q_h)^{\frac{3}{2}}}\frac{\tau^3}{3!} + \cdots$$

La limite est encore $-\frac{\tau^2}{2}$, et la loi réduite tend vers la loi de Gauss réduite lorsque $n_1, n_2, \ldots, n_k$ augmentent indéfiniment; mais pour rendre la démonstration du fait que les coefficients tendent tous vers zéro complètement rigoureuse, il est nécessaire d'admettre que la plus grande des quantités $n_h p_h q_h$ reste dans un rapport borné avec la plus petite.

26. Combinaison linéaire d'erreurs vérifiant la loi de Gauss. — Si l'erreur x commise dans la mesure d'une quantité Z vérifie la loi de Gauss, l'erreur quadratique moyenne étant m, la fonction caractéristique est

$$\Phi(t) = e^{-\frac{m^2 t^2}{2}}.$$

L'erreur qui résulte des conditions de l'expérience sur la quantité $U = \alpha Z$ est αx; elle vérifie aussi la loi de Gauss, avec la valeur $\alpha^2 m^2$ pour le carré moyen de l'erreur, la fonction caractéristique correspondante est

$$\Phi(t) = e^{-\frac{\alpha^2 m^2 t^2}{2}}.$$

Soit maintenant la combinaison linéaire

$$U = \alpha_1 Z_1 + \alpha_2 Z_2 + \ldots + \alpha_n Z_n \quad (3)$$

des quantités Z_i, soumises à des mesures dont les erreurs vérifient la loi de Gauss, avec les erreurs quadratiques moyennes

$$m_1, \quad m_2, \quad \ldots, \quad m_n.$$

L'erreur

$$u = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n$$

vérifie une loi dont la fonction caractéristique est le produit

$$e^{-\frac{\alpha_1^2 m_1^2 t^2}{2}} e^{-\frac{\alpha_2^2 m_2^2 t^2}{2}} \ldots e^{-\frac{\alpha_n^2 m_n^2 t^2}{2}} = e^{-\frac{t^2}{2} \sum \alpha_i^2 m_i^2}.$$

Cette loi est donc encore la loi de Gauss; l'erreur quadratique moyenne correspondante a pour valeur

$$m = \sqrt{\alpha_1^2 m_1^2 + \ldots + \alpha_n^2 m_n^2};$$

si k_1, k_2, ..., k_n et k sont les paramètres de précision correspondant aux déterminations de Z_1, Z_2, ..., Z_n et U, on a

$$\frac{1}{k^2} = \frac{\alpha_1^2}{k_1^2} + \frac{\alpha_2^2}{k_2^2} + \ldots + \frac{\alpha_n^2}{k_n^2}.$$

Cette propriété a été obtenue directement dans l'étude faite au Chapitre III du fascicule I pour la théorie des épreuves répétées. On peut l'étendre *d'une manière approximative*, et d'autant mieux que *les erreurs sont plus petites*, au cas d'une *fonction quelconque* des

quantités directement mesurées. Soit

$$U = f(X, Y, Z)$$

une fonction des quantités X, Y, Z; supposons que l'on connaisse de ces quantités des valeurs approchées, mesurées, égales à X_0, Y_0, Z_0, les erreurs $x = X - X_0$, $y = Y - Y_0$, $z = Z - Z_0$ vérifiant la loi de Gauss.

Si les erreurs sont petites, on a sensiblement pour

$$u = U - U_0 = f(X, Y, Z) - f(X_0, Y_0, Z_0)$$

le développement

$$u = x\frac{\partial f}{\partial X}(X_0, Y_0, Z_0) + y\frac{\partial f}{\partial Y}(X_0, Y_0, Z_0) + z\frac{\partial f}{\partial Z}(X_0, Y_0, Z_0),$$

donc u vérifie la loi de Gauss, l'erreur quadratique moyenne étant, si α, β, γ désignent les trois dérivées partielles ci-dessus,

$$m = \sqrt{\alpha^2 m_x^2 + \beta^2 m_y^2 + \gamma^2 m_z^2}$$

Cette propriété est fondamentale; on l'appelle souvent la propriété d'*invariance* ou de *stabilité* de la loi de Gauss.

27. Moyenne arithmetique des mesures. Poids des observations. Appliquons les résultats qui précèdent au cas où z_1, z_2, ..., z_n sont les résultats de n mesures de la même grandeur Z dans les mêmes conditions de précision. L'erreur commise en adoptant la valeur

$$z = \frac{z_1}{n} + \frac{z_2}{n} + \cdots + \frac{z_n}{n},$$

moyenne arithmétique des résultats, obéit à la loi de Gauss si les erreurs des mesures la vérifient de leur côté; si m est l'erreur quadratique moyenne d'une mesure, et μ celle de la moyenne arithmétique, on a

$$\mu^2 = n\left(\frac{m}{n}\right)^2,$$

donc

$$\mu = \frac{m}{\sqrt{n}};$$

si k est le paramètre de précision d'une mesure, celui de la moyenne arithmétique est donc $h = k\sqrt{n}$

La répétition des mesures augmente donc la précision du résultat, si l'on adopte comme valeur approchée la moyenne arithmétique des résultats directs. Il serait absurde, au point de vue pratique, d'en conclure la possibilité d'une amélioration indéfinie de la précision. En répétant un très grand nombre de fois une même expérience, en effet, on n'est jamais certain de n'avoir rigoureusement rien changé aux conditions dans lesquelles on opère.

Partant de cette idée bien naturelle qui consiste à adopter la moyenne arithmétique des résultats de plusieurs mesures, effectuées dans les mêmes conditions de précision, comme valeur approchée de la quantité à mesurer, nous pouvons toujours considérer un résultat obtenu par une mesure dont le paramètre de précision est k comme la moyenne arithmétique de n mesures dont le paramètre de précision serait $\frac{k}{\sqrt{n}}$.

Imaginons que l'on connaisse, pour une grandeur Z, le résultat z' d'une mesure ayant k' pour paramètre de précision, et le résultat z'' d'une mesure ayant k'' pour paramètre de précision. Si k'^2 et k''^2 sont proportionnels à deux nombres entiers n' et n'', on peut considérer z' et z'' comme les moyennes des résultats de n' et de n'' mesures ayant la même précision. La moyenne générale de ces $n' + n''$ mesures doit donc être adoptée comme valeur approchée, si l'on s'en tient à l'idée indiquée ci-dessus; cette valeur est

$$z = \frac{n'z' + n''z''}{n' + n''} = \frac{k'^2 z' + k''^2 z''}{k'^2 + k''^2}.$$

Etendant ce résultat au cas où k'^2 et k''^2 ne sont pas dans un rapport rationnel, et au cas où l'on possède n résultats $z_1, z_2, \ldots, z_n$ de mesures ayant pour paramètres de précision respectifs $k_1, k_2, \ldots, k_n$, on est conduit à la valeur approchée

$$z = \frac{k_1^2 z_1 + k_2^2 z_2 + \ldots + k_n^2 z_n}{k_1^2 + k_2^2 + \ldots + k_n^2}.$$

La forme de cette expression suggère l'emploi de coefficients, proportionnels à $k_1^2, k_2^2, \ldots, k_n^2$ appelés *poids* des observations. Il est

fait grand usage de cette notion dans les applications pratiques; on l'étend aux observations sur les lois d'erreur desquelles aucune hypothèse n'est faite, en définissant les poids comme *inversement proportionnels aux carrés moyens des erreurs*, dans le cas de la loi de Gauss ci-dessus, les deux définitions sont équivalentes.

CHAPITRE V.

LE PRINCIPE DE LA MOYENNE ARITHMÉTIQUE

28. Remarques générales sur la justification de la loi de Gauss — « La loi rigoureuse de probabilité des erreurs d'observation varie sans doute avec la grandeur mesurée, comme avec le choix de l'instrument et l'habileté de l'observateur ; elle est inaccessible aux géomètres. Euler, Bernoulli, Lagrange et Laplace ont fait des hypothèses démenties par les faits et mal justifiées par des preuves sans vraisemblance. Gauss, plus heureux, a déduit d'un raisonnement fort simple une loi que la démonstration laisserait douteuse, mais que les conséquences justifient. »

Ces quelques lignes, les premières du Chapitre consacré par Joseph Bertrand, dans son Traité classique, à la loi des erreurs d'observation, témoignent d'un scepticisme moins complet que celui de Poincaré, rapportant cette opinion de Lippmann : « Tout le monde croit à la loi de Gauss, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental »

La question de la loi des erreurs d'observation n'est pas du domaine des mathématiques pures Les efforts de Gauss pour la justifier au moyen d'un postulat purement mathématique, ont abouti à une démonstration que Joseph Bertrand qualifie de douteuse, et qui a fait l'objet de critiques sevères de sa part et de la part de Poincaré. C'est à l'étude du postulat de Gauss et de ses conséquences que nous consacrerons le présent Chapitre

Mais on peut considérer comme excessive l'opinion de Bertrand disant que la loi est inaccessible aux géomètres; dans la mesure où la loi de Gauss est vérifiée en effet par l'expérience, les mathématiques peuvent jouer leur rôle à son sujet comme à propos de toute question

de mathématiques appliquées. Comme en Physique mathématique, on peut utilement tenter de justifier des faits expérimentaux, qui ne peuvent être vérifiés que d'une manière approximative, à partir d'hypothèses de conception simple mais non vérifiables d'une manière directe. C'est à cet ordre d'idées que se rattachent les justifications basées sur la considération des *erreurs partielles*, et que nous étudierons aux Chapitres suivants.

29. La démonstration de Gauss. — En présence de plusieurs mesures inspirant la même confiance, tous les observateurs sont d'accord pour adopter la valeur z de la quantité mesurée Z égale à la *moyenne arithmétique*

$$z = \frac{z_1 + z_2 + \ldots + z_n}{n}$$

des résultats des mesures.

Tel est le postulatum sur lequel s'appuie la théorie appelée « première théorie de Gauss ».

On peut justifier la loi de Gauss, c'est-à-dire montrer que l'on a nécessairement

$$f(Z, z) = \frac{h}{\sqrt{\pi}} e^{-h^2(Z-z)^2},$$

moyennant les hypothèses suivantes :

1° La moyenne arithmétique des résultats est la valeur la plus probable de Z ;

2° Dans l'ignorance où l'on se trouve au sujet des probabilités *a priori* relatives aux valeurs de Z, on admet que $\varphi(Z)$ est une constante (Hypothèse de Bayes);

3° La fonction des probabilités élémentaires $f(Z, z)$ ne dépend que de l'erreur $x = Z - z$.

En effet, le problème consiste à déterminer la fonction $f(Z - z)$ de façon que la moyenne arithmétique

$$z = \frac{z_1 + z_2 + \ldots + z_n}{n}$$

assure, quels que soient $z_1, z_2, \ldots, z_n$, le maximum du produit

$$f(Z - z_1) f(Z - z_2) \ldots f(Z - z_n).$$

Posons

$$\omega(x_i) = \frac{f'(Z - z_i)}{f(Z - z_i)};$$

nous avons d'abord cette condition nécessaire, que la relation

$$\omega(x_1) + \omega(x_2) + \ldots + \omega(x_n) = 0$$

doit toujours être la conséquence de la relation

$$x_1 + x_2 + \ldots + x_n = 0$$

Si l'on admet que $\omega(x)$ possède une dérivée, on voit que

$$\omega'(x_1)\,dx_1 + \ldots + \omega'(x_n)\,dx_n = 0$$

doit toujours être la conséquence de

$$dx_1 + \ldots + dx_n = 0$$

il en résulte nécessairement

$$\omega'(x) = a$$

et

$$\omega(x) = ax$$

Pour $x = 0$, la quantité

$$\omega(x_1) + \ldots + \omega(x_n) = anx$$

doit passer du positif au négatif, afin que la probabilité soit bien maxima ; donc la constante a est nécessairement négative.

Nous avons ainsi, en posant $a = -2k^2$,

$$\text{Log} f(x) = -k^2x^2 + C$$

et

$$f(x) = A\,e^{-k^2x^2},$$

la condition

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

entraîne immédiatement que $A = \frac{k}{\sqrt{\pi}}$.

Cette justification est due à Gauss lui-même, elle est exposée dans son Mémoire *Theoria motus corporum celestium* (¹).

(¹) *Voir* la Note I de l'Ouvrage : *Méthode des moindres carrés*, par C.-F. Gauss, traduction J. Bertrand (Paris, Gauthier-Villars).

Il est facile de s'assurer qu'effectivement, si l'on admet la loi de Gauss et l'*hypothèse de Bayes* [$\varphi(Z) = \text{const.}$], la moyenne arithmétique des mesures est effectivement la valeur la plus probable. Car, vu les résultats des mesures faites, la probabilité *a posteriori* pour que Z soit compris entre les valeurs Z et $Z + dZ$ est

$$P(Z)\,dZ = \frac{e^{-k^2[(Z-z_1)^2+\cdots+(Z-z_n)^2]}\,dZ}{\int_{-\infty}^{+\infty} e^{-k^2[(Z-z_1)^2+\cdots+(Z-z_n)^2]}\,dZ},$$

le maximum de cette probabilité est réalisé par la valeur qui assure le minimum de la somme

$$(Z - z_1)^2 + \cdots + (Z - z_n)^2,$$

et cette valeur est bien la moyenne arithmétique.

30. Objections à la démonstration précédente. — « La fonction que nous venons de trouver, écrit Gauss, ne peut pas exprimer, en toute rigueur, la probabilité des erreurs, puisque, les erreurs possibles étant toujours renfermées entre certaines limites, la probabilité d'erreurs plus grandes devrait être toujours nulle, tandis que notre fonction a toujours une valeur finie. Cependant ce défaut, que présenterait également toute autre fonction analytique, n'a aucune importance dans les applications, parce que la valeur de notre fonction décroît si rapidement, pour peu que kx ait une valeur considérable, qu'on peut, en toute sûreté, la regarder alors comme équivalente à o. D'ailleurs, la nature de la question ne permettra jamais d'assigner les limites des erreurs avec une précision absolue. »

Ce langage montre que Gauss ne considérait pas sa théorie comme se trouvant à l'abri de toute critique. Des objections nombreuses lui ont, en effet, été faites.

L'une d'elles consiste à dire que l'interprétation mathématique adoptée pour le principe de la moyenne arithmétique, savoir que cette moyenne doit être la valeur *la plus probable*, n'est nullement imposée par la nature de la question; Joseph Bertrand exprime l'avis que l'interprétation d'après laquelle la moyenne arithmétique doit être la *valeur probable* est mieux justifiée.

Calculons la valeur probable de Z en admettant la loi de Gauss; ce

sera

$$\overline{Z} = \int_{-\infty}^{+\infty} Z\,P(Z)\,dZ.$$

Si l'on pose

$$z = \frac{z_1 + z_2 + \cdots + z_n}{n}, \qquad \zeta^2 = \frac{z_1^2 + z_2^2 + \cdots + z_n^2}{n},$$

l'exposant de la loi de Gauss s'écrit

$$-nk^2[Z^2 - 2zZ + \zeta^2]$$

si l'on divise haut et bas l'expression de $\overline{Z}$ par le facteur constant $e^{-nk^2(\zeta^2 - z^2)}$, il reste, en posant $Z = z + x$,

$$\overline{Z} = \frac{\int_{-\infty}^{+\infty} e^{-nk^2x^2}(z + x)\,dx}{\int_{-\infty}^{+\infty} e^{-nk^2x^2}\,dx}.$$

L'intégrale du numerateur se décompose en deux en separant les deux termes de la somme $z + x$, mais la deuxieme des integrales obtenues est nulle, et il reste bien $\overline{Z} = z$

Ainsi, en admettant la loi de Gauss, la moyenne arithmétique est à la fois la valeur la plus probable et la valeur probable Mais la reciproque n'est pas etablie, même en admettant les hypotheses 2° et 3° du paragraphe précédent, en ce qui concerne la valeur probable Du reste, ces hypothèses elles-mêmes sont contestables.

31. Discussion de Poincaré. — Poincaré a examine le problème de la détermination de toutes les lois de probabilite d'erreurs vérifiant la condition que la *valeur la plus probable*, ou bien la *valeur probable* de Z soit la moyenne arithmétique des résultats $z_1, z_2, \ldots, z_n$ et a étudié, pour chaque cas, la possibilité de s'affranchir des hypothèses 2° et 3° du § 29

Il détermine d'abord les fonctions $f(Z, z)$ et $\varphi(Z)$ du § 18 de manière que la moyenne arithmétique des valeurs $z_1, z_2, \ldots, z_n$ rende $P(Z)$ maximum, dans le cas le plus général.

En posant comme plus haut

$$\frac{f'_Z(Z, z)}{f(Z, z)} = \omega(Z - z),$$

la dérivée logarithmique

$$\omega(Z, z_1) + \ldots + \omega(Z, z_n) + \frac{\varphi'}{\varphi}$$

doit être nulle moyennant la condition

$$nZ = z_1 + z_2 + \ldots + z_n.$$

Pour tout système de valeurs des accroissements infiniment petits $dz_1, \ldots, dz_n$ ayant une somme nulle, à partir de valeurs fixes de $z_1, \ldots, z_n$ ayant pour somme nZ, on devra donc avoir

$$\omega'_{z_1} dz_1 + \ldots + \omega'_{z_n} dz_n = 0.$$

Cette condition entraîne $\omega'_z =$ const., la constante ne dépendant que de Z, donc

$$\omega(Z, z) = A'z + B',$$

$$\text{Log} f = Az + B + \psi(z),$$

et

$$f = \theta(z) e^{Az+B}$$

où A et B sont deux fonctions de Z.

Ces fonctions ne sont pas toutes deux arbitraires, car en portant la valeur $\omega = A'z + B'$ dans la dérivée logarithmique formée plus haut, nous obtenons

$$A'(z_1 + z_2 + \ldots + z_n) + nB' + \frac{\varphi'}{\varphi} = 0,$$

d'où, pour $z_1 + z_2 + \ldots + z_n = nZ$,

$$(A'Z + B')n + \frac{\varphi'}{\varphi} = 0,$$

condition qui doit être vérifiée quels que soient n et Z.

Cette condition entraîne :

1° Que $\frac{\varphi'}{\varphi} = 0$, donc que $\varphi =$ const. C'est l'hypothèse de Bayes, qui se trouve donc encore nécessaire dans les conditions de la discussion actuelle.

2° Que $A'Z + B' = 0$, ce qui conduit à l'expression

$$B = -\int A'Z\, dZ.$$

La loi des erreurs a donc en définitive pour fonction des probabi-

lités élémentaires

$$f(Z, z) = \theta(z)\, e^{Az - \int A' Z\, dZ}$$

où A désigne une fonction arbitraire de Z et θ une fonction arbitraire de z. Ce résultat est obtenu par la seule hypothèse que la moyenne arithmétique est la valeur la plus probable ; cette hypothese entraîne aussi, par ailleurs, l'hypothèse de Bayes

Si l'on admet, en outre, que $f(Z, z)$ ne dépend que de la différence $Z - z = x$, on se trouve dans les conditions du § 29, et la loi de probabilité doit se réduire à la loi de Gauss. Vérifions-le à partir de l'expression ci-dessus.

Posons $\operatorname{Log} \theta = \psi$, nous devons avoir l'identité

$$\frac{\partial}{\partial Z} \operatorname{Log} f = 0$$

une fois f exprimé en fonction de Z et x seuls Cette condition s'écrit

$$\psi' + A - x A' = 0$$

d'où, en dérivant par rapport à x, nous tirons $\psi'' + A' = 0$ Cette condition ne peut être vérifiée quels que soient x et Z que si l'on a simultanément

$$\psi'' = \text{const} = a, \qquad A' = \text{const} = -a$$

Intégrons, et tenons compte de l'identité

$$\psi' + A - x A' = 0,$$

nous avons

$$\psi' = a(Z - x) + \gamma, \qquad A = -aZ - \gamma,$$

donc

$$B = a\frac{Z^2}{2} + \beta$$

$$\psi = a\frac{(Z - x)^2}{2} + \gamma(Z - x)$$

et enfin

$$\operatorname{Log} f = \psi + A(Z - x) + B = a\frac{x^2}{2} + b$$

en posant $\beta + \gamma = b$. On retombe bien sur les résultats du § 29.

Poincaré examine ensuite la détermination des fonctions $f(Z, z)$ et $\varphi(Z)$ par la condition que la *valeur probable* de Z, à la suite des

observations ayant donné les résultats $z_1, z_2, \ldots, z_n$, soit la moyenne arithmétique.

Cette valeur probable est le rapport

$$\frac{\int_{-\infty}^{+\infty} Z f(Z, z_1) \ldots f(Z, z_n)\,\varphi(Z)\,dZ}{\int_{-\infty}^{+\infty} f(Z, z_1) \ldots f(Z, z_n)\,\varphi(Z)\,dZ}.$$

Or, la condition envisagée doit rester vraie même si les résultats z_1, $z_2, \ldots, z_n$ ont été donnés *chacun par p observations*, quelque grand que soit p. La valeur probable est alors donnée par le rapport des intégrales

$$\frac{\int_{-\infty}^{+\infty} Z\,\varphi(Z)\,F^p\,dZ}{\int_{-\infty}^{+\infty} \varphi(Z)\,F^p\,dZ},$$

où F désigne le produit

$$f(Z, z_1) \ldots f(Z, z_n).$$

Poincaré montre que la limite de ce rapport pour p infini est le rapport

$$\frac{\zeta\varphi(\zeta)}{\varphi(\zeta)} = \zeta$$

des éléments sous les signes d'intégration *pour la valeur ζ de Z qui rend* F *maximum*. Si cette valeur est la moyenne arithmétique, il faut donc que celle-ci assure le maximum du produit F, qui ne diffère du numérateur de P(Z) que par le facteur $\varphi(Z)$ qui caractérise la probabilité *a priori*.

Si l'on admet alors l'hypothèse de Bayes, on est ramené à l'analyse relative à la moyenne arithmétique considérée comme la valeur la plus probable.

Sinon, la discussion de Poincaré, pour le détail de laquelle nous renvoyons le lecteur au Chapitre X de son Traité, conduit à ce résultat singulier que la fonction $f(Z, z)$ a pour expression

$$f = \theta(z)\,e^{-\int \varphi(Z)(Z-z)\,dZ},$$

et dépend ainsi de la fonction $\varphi(Z)$. « Il n'y a aucune raison, écrit

Poincaré, pour que ces deux probabilités *a priori* dépendent l'une de l'autre La seule hypothèse raisonnable est donc de supposer $\varphi = 1$ pour retrouver la loi de Gauss. »

La discussion ne permet pas, en définitive d'affranchir le principe de la moyenne arithmétique, considérée comme devant être soit la *valeur la plus probable*, soit la *valeur probable*, ni de l'*hypothèse de Bayes* d'après laquelle toutes les valeurs de Z ont la même probabilité *a priori*, ni de celle d'après laquelle *la probabilité d'une erreur ne dépend que de cette erreur* $z - Z$.

Les critiques adressées à la démonstration de Gauss sont donc justifiées, et du reste Gauss, dans sa « deuxième théorie », a abandonné non seulement la démonstration, mais même la loi, se contentant d'établir sans elle les procédés mêmes de calcul auxquels elle conduit

Il convient de signaler ici que, comme l'a fait remarquer M Paul Lévy, la deuxième théorie de Gauss n'est qu'en apparence indépendante de la loi exponentielle des erreurs, puisque les hypothèses sur lesquelles s'appuie cette théorie sont celles qui permettent d'établir la dite loi d'une manière rigoureuse par d'autres méthodes que celle basée sur le principe de la moyenne arithmétique

CHAPITRE VI.

JUSTIFICATION DE LA LOI DE GAUSS PAR LA MÉTHODE DES MOMENTS

32. **La théorie des erreurs partielles**. — On peut envisager la question de la loi de probabilité des erreurs d'un point de vue tout à fait différent de celui adopté au Chapitre précédent, en faisant état de certaines hypothèses relatives à la genèse même des erreurs expérimentales.

Après Laplace et Bessel, de nombreux savants ont adopté l'hypothèse d'après laquelle l'erreur x est le résultat de la mise en jeu de nombreuses causes d'erreur, indépendantes, ayant chacune un effet minime. On appelle *erreurs partielles* les erreurs que donnerait sur la mesure chacune de ces causes, agissant seule, et l'on est fondé à penser que les lois auxquelles obéissent ces erreurs partielles sont symétriques, ou tout au moins que la valeur probable de chaque erreur partielle est nulle.

L'erreur globale est une fonction d'un très grand nombre de variables, qui sont les erreurs partielles. L'hypothèse suivante est généralement admise : *l'erreur globale est la somme algébrique des erreurs partielles.*

Les lois de probabilité des erreurs partielles étant inconnues, il est difficile d'obtenir un résultat précis et général en ce qui concerne la loi de probabilité de la somme d'un nombre fini de telles erreurs. Mais si l'on suppose ces dernières *extrêmement nombreuses* et *extrêmement petites*, on peut, en faisant certaines hypothèses très larges sur leurs lois de probabilité, démontrer rigoureusement que *la loi résultante est, à la limite, la loi de Gauss.*

Nous étudierons dans le présent Chapitre la justification obtenue, par la méthode de Tchebychef et de ses continuateurs, au moyen de la *théorie des moments.*

33. Indication fournie par la théorie des épreuves répétées — M. Émile Borel a formulé, au sujet des lois inconnues des probabilites des erreurs partielles, l'hypothèse très simple qui consiste à admettre que chaque cause d'erreur donne, si elle agit, une erreur déterminée en grandeur et en signe, et que, d'autre part, il y a pour chaque cause d'erreur une probabilité déterminée d'action. Toutes les erreurs partielles seraient donc des variables éventuelles d ordre un.

La simplification ainsi introduite peut s'expliquer de la maniere suivante : chaque cause d'erreur pouvant produire, suivant une certaine loi de probabilité, une erreur partielle, on peut décomposer le champ de variation de cette erreur partielle en intervalles assez petits pour que, dans chacun d'eux, l'erreur soit regardée comme constante On associe alors à chacune des erreurs élementaires constantes obtenues de cette manière une certaine probabilite de réalisation, et, moyennant des approximations dont l'effet est négligeable sur l'erreur globale eventuelle, on est conduit, vu le tres grand nombre d'erreurs partielles, à envisager plusieurs groupes de causes d'erreurs elementaires, comprenant n_1, n_2, ..., n_k causes, susceptibles de produire respectivement les erreurs α_1, α_2, ..., α_k avec les probabilités p_1, p_2, ..., p_k

Cette etude se rapportant aux seules erreurs accidentelles, nous devons admettre que, dans l'ensemble, les erreurs positives sont aussi probables que les erreurs négatives, ce qui se traduit par ce résultat que l'erreur résultante a une valeur moyenne

$$n_1\alpha_1p_1 + n_2\alpha_2p_2 + \dots + n_k\alpha_kp_k$$

égale à zéro.

Si donc chaque erreur partielle se produisait exactement suivant sa probabilité, l'erreur globale serait nulle, en réalité, les erreurs α_i se produisant $n_ip_i + x_i$ fois, l'erreur résultante a pour valeur

$$x = \alpha_1x_1 + \alpha_2x_2 + \dots + \alpha_kx_k.$$

Mais, quel que soit i, l'écart x_i obeit à la loi normale des écarts pour n_i assez grand, la valeur quadratique moyenne de x_i étant $m_i = \sqrt{n_ip_iq_i}$. L'erreur x, combinaison linéaire des variables x_1, x_2, ..., x_k obéissant à la loi de Gauss, vérifie elle-même cette loi, avec une valeur quadratique moyenne μ telle que

$$\mu^2 = \Sigma\alpha_i^2m_i^2 = \Sigma n_ip_iq_i\alpha_i^2.$$

Ainsi s'obtient une justification intuitive simple et intéressante de la loi de Gauss, il nous a paru utile de la résumer avant d'aborder les théories plus rigoureuses basées sur la méthode des moments et la méthode des fonctions caractéristiques.

34. Le problème des moments pour une loi de probabilité variable — Nous avons établi au § 11 qu'à la suite

$$c_0, \; c_1, \; c_2, \; \dots, \; c_{2n-1}$$

des $2n$ premiers moments d'une loi de probabilité dont la fonction des probabilités totales ne se réduit pas à une fonction à un nombre fini d'escaliers, on peut associer une fonction en escalier $G_n(x)$, qui tend uniformément vers $F(x)$ lorsque n augmente indéfiniment, si toutes les masses $\alpha_1, \alpha_2, \dots, \alpha_n$ tendent uniformément vers zéro. Cette condition est une *condition suffisante* pour que le problème général des moments relatifs à la fonction $F(x)$ *admette cette fonction pour solution unique.*

Imaginons que les moments dépendent d'un paramètre λ susceptible de prendre des valeurs formant un ensemble infini dénombrable ou continu, et supposons que, lorsque λ tend vers une limite déterminée l, les moments admettent des limites vérifiant *la condition d'unicité* ci-dessus.

Nous allons démontrer que *la fonction* $F(x, \lambda)$ *tend uniformément vers* $F(x, l)$.

Désignons en effet par $G_n(x, \lambda)$ la fonction à n escaliers à laquelle conduit le problème algébrique d'ordre n pour les valeurs $c_i(\lambda)$ des moments. Les quantités $\alpha_i(\lambda)$ et $x_i(\lambda)$ étant des fonctions algébriques des $c_i(\lambda)$, la fonction $G_n(x, \lambda)$ tend vers $G_n(x, l)$ lorsque λ tend vers l. Et si l'on considère un nombre ε aussi petit que l'on voudra, il est possible de prendre λ assez voisin de l, n étant donné, pour que, d'une part, on ait

$$|G_n(x, \lambda) - G_n(x, l)| < \varepsilon,$$

et que, d'autre part, les différences $|\alpha_i(\lambda) - \alpha_i(l)|$ soient inférieures à ε, le tout moyennant l'hypothèse $|\lambda - l| < \alpha_n$

Or, puisque les nombres $c_i(l)$ vérifient la condition d'unicité, on peut prendre n assez grand pour que tous les $\alpha_i(l)$ soient inférieurs à ε, c'est-à-dire que l'on ait, d'après le § 11,

$$|G_n(x, l) - F(x, l)| < \varepsilon;$$

alors les $\alpha_i(\lambda)$ seront moindres que 2ε, la fonction $G_n(x, \lambda)$ tendra uniformément vers la fonction $F(x, \lambda)$, de manière que

$$|F(x, \lambda) - G_n(x, \lambda)| < 2\varepsilon$$

Il s'ensuit, pour n supérieur à N, et $|\lambda - l|$ inférieur à la valeur σ_n correspondante, que nous avons, quel que soit x,

$$|F(x, \lambda) - F(x, l)| < 4\varepsilon$$

ce qui demontre le *théorème de convergence* énoncé plus haut.

35. Le problème général des moments pour la fonction de Gauss. — *Polynomes d'Hermite-Tchebychef.* — Si l'on envisage le problème algébrique d'ordre n pour la fonction

$$F(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-t^2} dt$$

le polynome $Q_n(x)$ correspondant a pour expression

$$Q_n(x) = \frac{e^{x^2}}{(-2)^n} \frac{d^n}{dx^n} e^{-x^2}$$

En effet, cette expression est bien un polynome d'ordre n dans lequel le coefficient de x^n est égal à 1, et d'autre part ce polynome vérifie bien la condition fondamentale d'orthogonalité

$$I = \int_{-\infty}^{+\infty} P(x) Q_n(x) e^{-x^2} dx = 0$$

pour tout polynome $P(x)$ de degré au plus égal à $n - 1$.

Pour le démontrer, on met l'intégrale I, au moyen de n intégrations par parties, sous la forme

$$\frac{1}{2^n} \int_{-\infty}^{+\infty} e^{-x^2} \frac{d^n}{dx^n} P(x) dx,$$

et l'on voit que $I = 0$ si le degré de $P(x)$ est inferieur à n.

Si $P(x)$ est de degré précisément égal à n, on a

$$\frac{d^n}{dx^n} P(x) = n! A_0$$

A_0 désignant le coefficient du terme en x^n, et il en résulte

$$1 = \frac{n!}{2^n} A_0 \sqrt{\pi}.$$

Les polynomes $Q_n(x)$ sont les *polynomes d'Hermite-Tchebychef*, déjà étudiés dans la Note II du fascicule I du présent Traité.

Ces polynomes fournissant les éléments de la fonction $G_n(x)$ associée à la loi de Gauss, nous allons examiner si les *conditions d'unicité* du § 11, rappelées ci-dessus au § 34, sont bien vérifiées dans le cas actuel ; elles consistent en ce que $a_1, a_2, \ldots, a_n$ tendent uniformément vers zéro pour n infini.

Soit à calculer par exemple a_1, d'après l'identité fondamentale (3) du § 10, nous avons, pour un polynome arbitraire $\theta(x)$ de degré $2n-1$ au plus,

$$a_1\theta(x_1) + \ldots + a_n\theta(x_n) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}\theta(x)e^{-x^2}dx.$$

Prenant

$$\theta(x) = \frac{Q_n Q_{n-1}}{x - x_1},$$

nous voyons que le premier membre se réduit à $a_1 Q'_n(x_1) Q_{n-1}(x)$; l'intégrale du second a pour valeur $\frac{(n-1)!}{2^{n-1}}$, d'après les propriétés des polynomes d'Hermite-Tchebychef, car $\frac{Q_n}{x-x_1}$ est un polynome d'ordre $n-1$ dans lequel le terme de plus haut degré a pour coefficient 1.

Dans ces conditions,

$$a_1 = \frac{(n-1)!}{2^{n-1}Q'_n(x_1)Q_{n-1}(x_1)}.$$

Or, l'identité évidente

$$e^{-x^2}Q_{n+1} = -\frac{1}{2}\frac{d}{dx}[e^{-x^2}Q_n],$$

mise sous la forme

$$Q_{n+1} = xQ_n - \frac{1}{2}Q'_n$$

montre que Q'_n possède la propriété d'orthogonalité

$$\int_{-\infty}^{+\infty} PQ'_n\,dx = 0$$

vis-à-vis de tout polynome P d'ordre $n-2$ au plus; Q_n ne diffère donc de Q_{n-1} que par un facteur constant, qui est certainement n. Nous en déduisons

$$a_1 = \frac{n!}{2^{n-1}\,Q'^2_n(x_1)}.$$

Il s'agit de montrer que a_1 tend vers zéro pour n infini. Cherchons une limite inférieure du facteur $Q'^2_n(x_1)$ qui figure au dénominateur; on peut l'écrire

$$f(x_1) = Q'^2_n(x_1) + \lambda\, Q^2_n(x_1).$$

Or, si nous calculons la dérivée de

$$f(x) = Q'^2_n(x) + \lambda\, Q^2_n(x),$$

nous avons

$$f'(x) = 2\,Q'_n(x)\,[Q''_n(x) + \lambda\, Q_n(x)].$$

Comme

$$Q'_n = n\,Q_{n-1}, \qquad Q''_n = n(n-1)\,Q_{n-2},$$

nous pouvons écrire

$$f'(x) = 2\,Q'_n[n(n-1)Q_{n-2} + \lambda\, Q_n].$$

Prenons $\lambda = 2n$. La dérivée se réduit à

$$f'(x) = 4n\,Q_n\,Q'_n + 2n(n-1)\,Q'_n\,Q_{n-2}$$

et, puisque

$$Q_n = x\,Q_{n-1} - \frac{n-1}{2}\,Q_{n-2},$$

ainsi que le montre la combinaison des relations

$$Q_n = x\,Q_{n-1} - \frac{1}{2}\,Q'_{n-1} \qquad \text{et} \qquad Q'_{n-1} = (n-1)\,Q_{n-2},$$

nous écrivons cette dérivée

$$f'(x) = 4nx\,Q'_n\,Q_{n-1} = 4x\,Q'^2_n(x).$$

Il suit de là que *le minimum de $f(x)$ est $f(0)$*, donc, qu'en définitive, quel que soit $i = 1, 2, \ldots, n$, a_i est moindre que

$$A_i = \frac{n!}{2^{n-1}[Q'^2_n(0) + 2n\,Q^2_n(0)]}.$$

L'expression entre crochets, qu'on écrit

$$n^2\,Q^2_{n-1}(0) + 2n\,Q^2_n(0),$$

a deux expressions différentes suivant la parité de n.

Cas de n pair. — Soit $n = 2p$. Alors $Q_{n-1}(0)$ est nul, et l'expression se réduit au terme

$$2n\,Q_n^2(0) = 4p\,Q_{2p}^2(0).$$

Or, la relation de récurrence

$$Q_n = x\,Q_{n-1} - \frac{n-1}{2}\,Q_{n-2}$$

donne, pour $x = 0$, $n = 2p$,

$$Q_{2p}(0) = -\frac{2p-1}{2}\,Q_{2p-2}(0),$$

d'où, de proche en proche, la valeur

$$Q_{2p}(0) = (-1)^p\,\frac{1.3.5\ldots 2p-1}{2^p}.$$

Pour $n = 2p$, nous avons donc

$$A_i = \frac{2p!}{2p\,2^{2p}\,\dfrac{[1.3.5\ldots 2p-1]^2}{2^{2p}}} = \frac{2.4.6\ldots 2p-2}{1.3.5\ldots 2p-1},$$

et, d'après la formule de Wallis, suivant laquelle

$$\frac{(2.4.6\ldots 2p)^2}{(1.3.5\ldots 2p-1)^2\,(2p+1)}$$

tend vers $\frac{\pi}{2}$, nous voyons que A_i, pour p infini, est équivalent à

$$\frac{1}{\sqrt{2p-1}}\sqrt{\frac{\pi}{2}};$$

il tend effectivement vers zéro.

Cas de n impair. — Soit $n = 2p+1$. Alors $Q_n(0)$ est nul, et l'expression entre crochets se réduit au terme $(2p+1)^2\,Q_{2p}^2(0)$, on a par conséquent

$$A_i = \frac{(2p+1)!}{2^{2p}\,(2p+1)^2\,\dfrac{[1.3.5\ldots 2p-1]^2}{2^{2p}}} = \frac{2.4.6\ldots 2p}{1.3.5\ldots 2p+1} \sim \frac{1}{\sqrt{2p+1}}\sqrt{\frac{\pi}{2}}.$$

En somme, pour n tendant vers l'infini, tous les a_i sont moindres qu'une même fonction de n, laquelle est un infiniment petit équi-

valent à $\sqrt{\frac{\pi}{2n}}$. Ils tendent donc uniformément vers zero, et les conditions d'unicité sont vérifiées.

36. Théorème limite fondamental — La loi de Gauss vérifiant les conditions d'unicite, il résulte du théoreme de convergence du § 34 que si les moments d'une loi variable de probabilite à un paramètre tendent, dans certaines conditions, vers les moments

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} t^r e^{-t^2}\,dt$$

de la loi de Gauss, la fonction des probabilités totales correspondantes tend vers la fonction

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-t^2}\,dt$$

Le théorème suivant, énonce par Bienaymé (1853) et démontre rigoureusement par Tchebychef, permet, dans ces conditions, d'affirmer que la loi de probabilité relative à la somme d'un très grand nombre de petites erreurs partielles tend vers la loi de Gauss, à la limite, moyennant certaines hypotheses très larges sur les lois composantes

THÉORÈME. — *Si $X = X_1 + X_2 + \dots + X_n$ est la somme de n variables éventuelles indépendantes appartenant à une suite illimitée, et si μ désigne la valeur quadratique moyenne de X, les valeurs moyennes des quantités $\left(\frac{X}{\mu\sqrt{2}}\right)^r$ tendent, quel que soit r, vers les moments*

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} x^r e^{-x^2}\,dx$$

correspondant à la loi de Gauss.

Les hypothèses, précisées par Liapounoff, sont les suivantes .

1° Les valeurs moyennes des quantités X_i sont toutes nulles;
2° Leurs valeurs quadratiques moyennes m_i sont toutes finies;
3° Les valeurs moyennes μ_i^r des quantités $|X_i|^r$ sont finies quel

que soit r, et les rapports

$$\frac{\mu_1^{(r)} + \mu_2^{(r)} + \dots + \mu_n^{(r)}}{(m_1 + m_2 + \dots + m_n)^{\frac{r}{2}}}$$

tendent vers zéro pour n infini, quel que soit le nombre entier r supérieur à 2.

Nous désignerons par $m_i^{(r)}$ la valeur moyenne de X_i^r. Il est clair que nous avons $\mu_i^{(r)} \geqq |m_i^{(r)}|$, le cas de l'égalité étant réalisé pour r pair.

Nous allons d'abord démontrer que *tous les moments d'ordre impair* $M(X^{2k+1})$ *tendent vers zéro*.

Nous avons, quelle que soit la parité de r,

$$(X_1 + X_2 + \dots + X_n)^r = \sum \frac{r!}{\alpha!\,\beta!\dots\lambda!} S_{\alpha\beta\dots\lambda},$$

où $S_{\alpha\beta\dots\lambda}$ désigne la fonction symétrique $\sum X_1^\alpha X_2^\beta \dots X_n^\lambda$ obtenue, pour chaque système de nombres entiers α, β, ..., λ ayant pour somme r, en permutant arbitrairement les indices inférieurs. Il s'ensuit que

$$M(X^r) = \sum \frac{r!}{\alpha!\,\beta!\dots\lambda!} T_{\alpha\beta\dots\lambda},$$

$T_{\alpha\beta\dots\lambda}$ désignant la même fonction symétrique en $m_1^{(\alpha)}, m_2^{(\beta)}, \dots, m_n^{(\lambda)}$.

Tous les termes de cette fonction dans lesquels l'un au moins des exposants α, β, ..., λ est égal à 1 sont nuls, d'après la première condition de Liapounoff.

S'il n'y a pas d'exposant égal à 1, on a certainement

$$|m_1^{(\alpha)} m_2^{(\beta)} \dots m_n^{(\lambda)}| \leqq \mu_1^{(\alpha)} \mu_2^{(\beta)} \dots \mu_n^{(\lambda)},$$

et

$$|T_{\alpha\beta\dots\lambda}| \leqq [\mu_1^{(\alpha)} + \mu_2^{(\alpha)} + \dots + \mu_n^{(\alpha)}] \dots [\mu_1^{(\lambda)} + \mu_2^{(\lambda)} + \dots + \mu_n^{(\lambda)}];$$

si donc on considère la quantité

$$M\left(\frac{X}{\mu\sqrt{2}}\right)^r = \frac{M(X^r)}{2^{\frac{r}{2}}(m_1 + m_2 + \dots + m_n)^{\frac{r}{2}}},$$

le numérateur est formé : 1° de termes nuls comme ayant au moins un des exposants α, β, ..., λ égal à 1 ; 2° de termes majorés par les

produits

$$[\mu_1^{(\alpha)}+\mu_2^{(\alpha)}+\ldots+\mu_n^{(\alpha)}]\ldots[\mu_1^{(\lambda)}+\mu_2^{(\lambda)}+\ldots+\mu_n^{(\lambda)}]$$

correspondants. La quantité $M\left(\frac{X}{\mu\sqrt{2}}\right)^r$ est donc inférieure en valeur absolue à la somme des produits

$$\frac{\mu_1^{(\alpha)}+\mu_2^{(\alpha)}+\ldots+\mu_n^{(\alpha)}}{2^{\frac{\alpha}{2}}(m_1+m_2+\ldots+m_n)^{\frac{\alpha}{2}}}\cdots\frac{\mu_1^{(\lambda)}+\mu_2^{(\lambda)}+\ldots+\mu_n^{(\lambda)}}{2^{\frac{\lambda}{2}}(m_1+m_2+\ldots+m_n)^{\frac{\lambda}{2}}},$$

pour tous les groupes possibles d'exposants $\alpha, \beta, \ldots, \lambda$ tous différents de 1 et ayant pour somme r.

Si r est impair, l'un au moins, soit δ, des exposants $\alpha, \beta, \ldots, \lambda$ est impair et au moins égal à 3, le rapport

$$\frac{\mu_1^{(\delta)}+\mu_2^{(\delta)}+\ldots+\mu_n^{(\delta)}}{(m_1+m_2+\ldots+m_n)^{\frac{\delta}{2}}}$$

correspondant tend vers zéro, d'après la troisième condition de Liapounoff. Donc tous les produits précédents tendent vers zéro, et tous les moments d'ordre impair tendent bien vers ceux de la loi de Gauss, qui sont nuls.

Soit maintenant à démontrer que, pour $r=2p$, la valeur moyenne de $\left(\frac{X}{\mu\sqrt{2}}\right)^r$ tend vers $\frac{1\cdot 3\cdot 5\ldots(2p-1)}{2^p}$ (§ 22).

D'après l'étude qui précède et la troisième condition de Liapounoff, nous n'avons à tenir compte, dans le calcul de la somme

$$\Sigma\frac{r!}{\alpha!\,\beta!\ldots\lambda!}T_{\alpha\beta\ldots\lambda},$$

que des termes $T_{\alpha\beta\ldots\lambda}$ pour lesquels p exposants sont égaux à 2, et p à zéro. Les coefficients sont tous égaux à $\frac{2p!}{2^p}$, et nous avons

$$\lim M\left[\frac{X_1+X_2+\ldots+X_n}{\sqrt{2(m_1+m_2+\ldots+m_n)}}\right]^{2p}=\frac{2p!}{2^{2p}}\lim\Sigma\frac{m_1m_2\ldots m_p}{(m_1+m_2+\ldots+m_n)^p}.$$

Or, par le développement de $(m_1+m_2+\ldots+m_n)^p$, nous avons

$$1=p!\,\Sigma\frac{m_1m_2\ldots m_p}{(m_1+m_2+\ldots+m_n)^p}+\Sigma\frac{p!}{h!\,k!\ldots l!}\,\frac{m_1^h m_2^k\ldots m_n^l}{(m_1+m_2+\ldots+m_n)^p},$$

en désignant par h, k, ..., l des exposants dont l'un au moins est supérieur à 1, et dont la somme est égale à p. Chaque terme de la deuxième partie du second membre est inférieur ou égal au terme

$$\frac{p!}{h!\,k!\dots l!}\,\frac{\mu_1^{2h}\mu_2^{2k}\dots\mu_n^{2l}}{(m_1+m_2+\dots+m_n)^p}$$

correspondant, car on établit aisément, par récurrence, l'inégalité

$$m_i^l \leqq \mu_i^{2l}$$

pour $i = 1, 2, \dots, n$ et l entier supérieur ou égal à 1.

Dans ces conditions, la fonction symétrique obtenue, pour h, k, ..., l donnés, en prenant pour les indices inférieurs toutes les permutations possibles des nombres 1, 2, ..., n, est inférieure au produit

$$\frac{p!}{h!\,k!\dots l!}\,\frac{\mu_1^{2h}+\mu_2^{2h}+\dots+\mu_n^{2h}}{(m_1+m_2+\dots+m_n)^h}\cdots\frac{\mu_1^{2l}+\mu_2^{2l}+\dots+\mu_n^{2l}}{(m_1+m_2+\dots+m_n)^l},$$

qui comprend au moins un facteur tendant vers zéro pour n infini. En définitive, $\sum\frac{m_1 m_2\dots m_p}{(m_1+m_2+\dots+m_n)^p}$ tend vers $\frac{1}{p!}$, et la limite de la valeur moyenne de $\left(\frac{\Lambda}{\mu\sqrt{2}}\right)^{2p}$ est bien $\frac{1\cdot 3\cdot 5\dots(2p-1)}{2^p}$.

CHAPITRE VII.

JUSTIFICATION DE LA LOI DE GAUSS PAR LA MÉTHODE DES FONCTIONS CARACTÉRISTIQUES.

37. Fonction caractéristique d'une loi de probabilité variable. — Le but du présent Chapitre est d'indiquer les grandes lignes de la justification de la loi de Gauss donnée par M. Paul Lévy au moyen de la considération des fonctions caractéristiques. Nous nous bornerons à l'essentiel, et nous renverrons le lecteur, pour plus de détails, au Traité de M. Paul Lévy [1].

Nous savons, d'après le § 13, qu'une loi de probabilité est entièrement déterminée par la connaissance de sa fonction caractéristique, la fonction des probabilités totales étant donnée par la relation

$$F(X) - F(0) = \frac{1}{2p} \, vp \int_{-\infty}^{+\infty} \frac{1 - e^{-itX}}{it} \Phi(t) \, dt$$

Considérons une famille de lois de probabilité dépendant d'un paramètre λ; nous allons démontrer une *propriété de convergence* qui n'est pas sans analogie avec celle du § 34 : *si la fonction caractéristique* $\Phi(t, \lambda)$ *tend uniformément, dans tout intervalle fini* $(-C, +C)$, *vers sa limite* $\Phi(t, l)$, *la fonction des probabilités totales* $F(X, \lambda)$ *tend uniformément vers sa limite* $F(X, l)$.

D'après le § 13, l'intégrale

$$J(C, \lambda) = \int_{-C}^{+C} \Phi(t, \lambda) \frac{1 - e^{-itX}}{it} \, dt$$

qu'on écrit aussi

$$\int_{-\infty}^{+\infty} 2 \frac{\sin C y}{y} [F(y + X, \lambda) - F(y, \lambda)] \, dy,$$

---

(1) Paul Lévy, *Calcul des Probabilités*, Paris, 1925 (Gauthier-Villars).

a pour limite

$$2\pi[F(X, \lambda) - F(0, \lambda)]$$

lorsque C tend vers $+\infty$

Il suffit de démontrer que, quel que soit λ dans le voisinage de l, la convergence de $J(C, \lambda)$ vers cette limite *est uniforme par rapport à λ*; car les trois parties, suivant lesquelles on peut décomposer

$$J(\infty, l) - J(\infty, \lambda)$$

en

$$[J(\infty, l) - J(C, l)] + [J(C, l) - J(C, \lambda)] + [J(C, \lambda) - J(\infty, \lambda)],$$

pourront alors être rendues moindres que tout nombre donné, les deux parties extrêmes en prenant C assez grand et la partie moyenne en prenant λ assez voisin de l

Portons notre attention sur la différence

$$\Delta = \int_{-\infty}^{+\infty} \frac{2 \sin C y}{y} F(y + X, \lambda) dy - 2\pi F(X, \lambda)$$

des termes de $J(C, \lambda)$ et $J(\infty, \lambda)$ qui correspondent à la valeur X de la variable. D'après le résultat élémentaire

$$\int_{-\infty}^{+\infty} \frac{\sin C y}{y} dy = \pi \qquad \text{(pour C positif)},$$

nous pouvons écrire cette différence

$$\Delta = \int_{-\infty}^{+\infty} \frac{2 \sin C y}{y} [F(y + X, \lambda) - F(X, \lambda)] dy,$$

ou, en posant $x = Cy$,

$$\Delta = \int_{-\infty}^{+\infty} \frac{2 \sin x}{x} \left[F\left(X + \frac{x}{C}, \lambda\right) - F(X, \lambda)\right] dx$$

Il s'agit de montrer que, quel que soit λ, on peut rendre Δ moindre que tout nombre donné ε. Écrivons Δ sous la forme

$$\Delta_1 + \Delta_2 + \Delta_3 = \int_{-\infty}^{-n\pi} + \int_{-n\pi}^{+n\pi} + \int_{n\pi}^{+\infty},$$

nous allons d'abord prendre n assez grand pour que Δ_1 et Δ_3 soient moindres que $\frac{\varepsilon}{3}$.

Considérons par exemple Δ_1 ; cette quantité se présente comme une différence de deux séries alternées, dont les termes décroissent constamment, puisque la fonction F décroît jusqu'à zéro lorsque X varie de $-n\pi$ à $-\infty$. Chacune de ces séries a un module moindre que celui du premier terme

$$\int_{-(n-1)\pi}^{-n\pi} \left| \frac{2 \sin t}{t} F \right| dt$$

lui-même inférieur à $\frac{4}{n\pi}$, ainsi Δ_1 est inférieur en module à $\frac{8}{n\pi}$, donc à $\frac{\varepsilon}{3}$ dès que n dépasse $\frac{24}{\pi\varepsilon}$.

Une démonstration toute pareille, en écrivant le crochet

$$\left[[1 - F(X, \lambda)] - \left[1 - F\left(X - \frac{t}{C}, \lambda\right) \right] \right]$$

montre que, pour $n > \frac{24}{\pi\varepsilon}$, Δ_3 est inférieur à $\frac{\varepsilon}{3}$.

Le nombre n une fois choisi, nous allons montrer qu'il est possible de prendre C assez grand pour que Δ_2 soit lui aussi inférieur à $\frac{\varepsilon}{3}$. Supposons qu'il existe une fonction $f(X, \lambda)$ des probabilités élémentaires, et que cette fonction soit moindre qu'un nombre fini M, alors nous avons

$$F\left(X - \frac{t}{C}, \lambda\right) - F(X, \lambda) = \int_X^{X - \frac{t}{C}} f(X, \lambda) \, dX < M \frac{t}{C}$$

et

$$|\Delta_2| < \int_{-n\pi}^{-n\pi} \frac{2M}{C} |\sin t| \, dt \quad \text{ou} \quad \frac{8nM}{C},$$

on pourra rendre $|\Delta_2|$ moindre que $\frac{\varepsilon}{3}$ pour C assez grand, n étant donné.

Après avoir ainsi utilisé l'hypothèse d'une fonction des probabilités élémentaires bornée, M. Paul Lévy montre que cette hypothèse n'est pas indispensable, et que le théorème de convergence est général.

38. **Loi résultante réduite.** — Nous avons étudié au § 24 la réduction d'une loi de probabilité, et démontré notamment la

relation

$$\overline{\Phi}(\tau) = e^{-\frac{\mu_1}{m} i\tau}\,\Phi\left(\frac{\tau}{m}\right)$$

entre les fonctions caractéristiques d'une loi réduite et non réduite.

Formons le logarithme $\overline{\psi}(\tau)$ de $\overline{\Phi}(\tau)$; nous avons

$$\overline{\psi}(\tau) = -\frac{i\tau}{m}\mu_1 + \operatorname{Log}\left(1 + i\mu_1\frac{\tau}{m} - \frac{\mu_2}{2}\frac{\tau^2}{m^2} + \cdots\right)$$

et nous voyons que $\overline{\psi}(\tau)$ est, par rapport à τ, un infiniment petit du second ordre, de partie principale

$$\frac{\tau^2}{2}\left(-\frac{\mu_2}{m^2} + \frac{\mu_1^2}{m^2}\right),$$

ou $-\frac{\tau^2}{2}$, puisque

$$m^2 = \mu_2 - \mu_1^2.$$

On peut donc écrire

$$\overline{\psi}(\tau) = -\frac{\tau^2}{2}[1 + \omega(\tau)],$$

$\omega(\tau)$ désignant une fonction de τ qui tend vers zéro avec τ. *Dans le cas de la loi de Gauss, les calculs du* § 24 *montrent que la fonction* $\omega(\tau)$ *est identiquement nulle.*

Considérons la variable éventuelle

$$x = x_1 + x_2 + \cdots + x_n$$

somme d'un grand nombre de variables, qui sont des erreurs partielles obéissant à des lois de probabilité déterminées, sur lesquelles *nous ferons les hypothèses que les valeurs moyennes* $M(x_i)$ *sont toutes nulles, et les valeurs quadratiques moyennes*

$$m_i = \sqrt{M(x_i^2)}$$

toutes finies.

Dans ces conditions, les lois de probabilité composantes et la loi résultante peuvent être toutes réduites par les changements d'unité

$$x_i = m_i\xi_i, \qquad x = m\xi,$$

avec

$$m^2 = m_1^2 + m_2^2 + \cdots + m_n^2.$$

Nous avons, en ce qui concerne les lois non réduites,

$$\Phi_i(t) = 1 - \frac{m_i^2}{2} t^2$$

$$\Phi(t) = 1 - \frac{m^2}{2} t^2$$

et

$$\text{Log } \Phi_i(t) = \psi_i(t) = - \frac{m_i^2 t^2}{2} [1 - \omega_i(m_i t)]$$

$$\text{Log } \Phi(t) = \psi(t) = - \frac{m^2 t^2}{2} [1 - \omega(mt)]$$

avec la relation suivante, conséquence de la composition des lois de probabilité par multiplication des fonctions caractéristiques, donc addition des logarithmes $\psi(\tau)$

$$m^2 \omega(mt) = \sum_1^n m_i^2 \omega_i(m_i t)$$

Avec la variable $\tau = mt$, $\omega(\tau)$ désignant la fonction mise en évidence ci-dessus dans l'expression du logarithme $\overline{\psi}(\tau)$ de la fonction caractéristique de la loi réduite, nous avons donc

$$\omega(\tau) = \sum_1^n \frac{m_i^2}{m^2} \omega_i\left(\frac{m_i}{m} \tau\right)$$

39. **Limite de la loi résultante réduite.** — Pour montrer que la loi résultante réduite tend vers la loi de Gauss réduite il suffit de montrer que $\overline{\psi}(\tau)$ tend uniformément, dans tout intervalle fini $(-C\ +C)$, vers $-\frac{\tau^2}{2}$, *c'est-à-dire que* $\omega(\tau)$ *tend vers zéro.*

M Paul Lévy s'appuie sur deux hypothèses fondamentales .

1° Les fonctions $\omega_i(\tau)$ sont toutes majorées, dans un certain intervalle $(-\alpha, +\alpha)$, par une même fonction $\Omega(\tau)$, nulle pour $\tau = 0$,

2° Le nombre n des lois composantes augmentant indéfiniment, le rapport $\frac{l}{m}$ tend vers zéro, l désignant le plus grand des m_i

Soit, dans ces conditions, un nombre ε donné aussi petit que l'on voudra, nous allons montrer qu'on peut prendre n assez grand pour

que $|\omega(\tau)|$ soit inférieur à ε, quel que soit τ dans un intervalle donné $(-C, +C)$.

D'après les hypothèses faites, tous les rapports $\frac{m_l}{m}$ sont en effet, à partir d'une certaine valeur de n, moindres que $\frac{\alpha}{C}$. Toutes les quantités $\frac{m_l}{m}\tau$ sont alors comprises entre $-\alpha$ et $+\alpha$, si τ est compris entre $-C$ et $+C$, et toutes les fonctions $\omega_l\left(\frac{m_l}{m}\tau\right)$ sont inférieures en module à $\Omega\left(\frac{m_l}{m}\tau\right)$.

Nous pouvons, sans rien ajouter à la première hypothèse qui soit essentiel, supposer la fonction Ω paire et croissante avec $|\tau|$. A partir de n assez grand, nous aurons, quel que soit l,

$$\left|\omega_l\left(\frac{m_l}{m}\tau\right)\right| < \Omega\left(\frac{l}{m}C\right),$$

et, le second membre tendant vers zéro avec la variable $\frac{lC}{m}$, on peut le rendre inférieur à ε en prenant n assez grand, car alors $\frac{l}{m}$ tend vers zéro.

La somme Σm_l^2 étant égale à m^2, nous voyons que nous avons, quel que soit τ dans l'intervalle $-C, +C$,

$$-\varepsilon < \omega(\tau) < \varepsilon,$$

le rapport

$$\frac{\Phi(\tau)}{e^{-\frac{\tau^2}{2}}}$$

est compris entre $e^{-\varepsilon}$ et e^{ε}, et la différence $\Phi(\tau) - e^{-\frac{\tau^2}{2}}$ comprise entre $e^{\varepsilon} - 1$ et $e^{-\varepsilon} - 1$.

Ainsi, d'après le théorème de convergence du § 37, la *loi résultante réduite tend vers la loi de Gauss réduite*

40. Discussion des hypothèses faites — Examinons sur quelles hypothèses s'appuient les justifications que nous venons d'étudier au Chapitre VI et au Chapitre VII. Ces justifications doivent être considérées comme d'autant mieux applicables à la réalité que les hypothèses sont plus naturelles vis-à-vis de cette réalité

Il y a d'abord une hypothèse fondamentale sur la genèse de l'erreur

globale à partir des causes élémentaires d'erreur, ensuite diverses hypothèses sur les lois de probabilité des erreurs partielles.

L'hypothèse fondamentale, commune aux deux théories que nous avons étudiées, est celle de *l'additivité des erreurs partielles*, qui sont en outre considérées comme *indépendantes*. Cette hypothèse paraît bien naturelle, mais on peut en contester cependant le bien-fondé ; signalons que M. Fréchet, en la remplaçant par celle d'après laquelle « l'erreur globale est égale à la plus grande des erreurs partielles », a obtenu une loi de probabilité différente de la loi de Gauss. Cette nouvelle hypothèse n'est sans doute pas plus naturelle que celle de l'additivité, et du reste M. Fréchet ne prétend pas qu'elle le soit. Nous renvoyons le lecteur, pour cette critique des démonstrations de la loi de Gauss, à l'exposé de M. Fréchet dans le fascicule III du Tome I du présent Traité [1].

Examinons maintenant les hypothèses relatives aux lois de probabilité des erreurs partielles.

Les deux justifications admettent d'abord que la valeur moyenne de chaque erreur partielle est nulle ; cette hypothèse est beaucoup moins précise que celle énoncée par Poincaré, selon laquelle les lois de probabilités composantes sont toutes symétriques ; elle traduit d'une manière très plausible l'influence du hasard sur chaque erreur accidentelle élémentaire.

Les deux justifications admettent aussi que les valeurs quadratiques moyennes des erreurs partielles sont toutes finies. Cette hypothèse précise l'idée fondamentale selon laquelle les grandes valeurs d'une erreur accidentelle élémentaire sont très peu probables.

A partir de ce moment, les hypothèses envisagées concernent, non plus chaque loi composante, mais les lois composantes dans leur ensemble ; et ces hypothèses ne sont pas les mêmes dans les deux théories.

Dans la théorie de Tchebychef-Liapounoff, il s'agit des sommes des valeurs moyennes des puissances successives des erreurs absolues élémentaires ; ces sommes sont supposées, à la limite, infiniment petites par rapport aux puissances de même rang de l'erreur quadra-

[1] *Voir* aussi M. Fréchet, *Sur l'hypothèse de l'additivité des erreurs partielles* (*Bulletin des Sciences mathématiques*, 1928) — *Sur la loi de probabilité de l'écart maximum* (*Annales de la Société mathématique polonaise*, 1928).

tique moyenne globale, l'idée traduite par cette troisième condition de Liapounoff est encore celle de la petite probabilité de grandes valeurs d'une portion importante de l'ensemble des erreurs partielles

Dans la théorie de M Paul Lévy, il y a deux hypothèses distinctes, dont l'une, celle qui concerne les fonctions ω_i, paraît *a priori* bien abstraite; mais l'auteur montre que cette hypothèse revient à dire que, pour chaque loi composante réduite, la convergence des intégrales

$$\int_{-\infty}^{+\infty} \xi^2 \, d\bar{F}(\xi)$$

vers leur limite 1 est uniforme par rapport au paramètre définissant cette loi. La deuxième, selon laquelle le rapport $\frac{l}{m}$ tend vers zéro, exprime l'idée qu'aucune des erreurs partielles ne fournit à elle seule une partie notable de l'erreur globale Dans leur ensemble, elles sont bien en harmonie avec la notion, admise *a priori*, d'une erreur globale, somme d'un très grand nombre d'erreurs partielles indépendantes du même ordre de grandeur.

CHAPITRE VIII.

COMBINAISON DES OBSERVATIONS VÉRIFIANT LA LOI DE GAUSS

I — COMBINAISON D'OBSERVATIONS DIRECTES

41. **Observations directes d'égale précision.** — Le problème que nous allons étudier a été énoncé au § 18 : n déterminations expérimentales d'une même grandeur Z ayant donné les résultats z_1, z_2, ..., z_n, quelle valeur convient-il d'adopter pour Z, et quelle erreur peut-on craindre sur la valeur ainsi adoptée?

Si l'on admet que ces déterminations expérimentales vérifient la loi de Gauss, le paramètre de précision étant le même pour toutes nous avons montré au § 29 que, moyennant l'hypothèse de Bayes, la *valeur la plus probable* de Z est la moyenne arithmétique

$$z = \frac{z_1 + z_2 + \ldots + z_n}{n},$$

qui assure le minimum de la somme

$$(Z - z_1)^2 + (Z - z_2)^2 + \ldots + (Z - z_n)^2$$

des carrés des différences appelées *résidus* des observations. Nous avons même constaté au § 30 que la moyenne arithmétique z est également *la valeur probable* de Z. Ces deux interprétations, selon lesquelles la valeur la plus plausible est la valeur la plus probable, ou la valeur probable, conduisent donc à la même solution z, que nous adopterons par conséquent.

Quant à l'erreur à craindre sur la valeur z ainsi adoptée, nous pouvons la caractériser de la manière suivante. Nous admettons que les erreurs

$$x_1 = Z - z_1, \quad \ldots, \quad x_n = Z - z_n$$

obéissent toutes à la loi de Gauss avec le même paramètre de précision k, *d'ailleurs inconnu*.

Soient $y_1 = z - z_1, \ldots, y_n = z - z_n$ les *résidus*, c'est-à-dire les différences entre la moyenne arithmétique et les résultats d'observation, et soit $x = Z - z$ l'erreur commise en adoptant la valeur z. Nous avons les relations

$$\begin{aligned} y_1 &= x_1 - x, \\ y_2 &= x_2 - x, \\ &\ldots \\ y_n &= x_n - x, \\ y_1 + y_2 + &\cdots + y_n = 0. \end{aligned}$$

Ajoutons-les membre à membre après multiplication par $\frac{n-1}{n}$, $-\frac{1}{n}, \ldots, -\frac{1}{n}, \frac{1}{n}$; nous obtenons

$$y_1 = \frac{n-1}{n} x_1 - \frac{1}{n} x_2 - \cdots - \frac{1}{n} x_n,$$

et, de même,

$$y_2 = -\frac{1}{n} x_1 + \frac{n-1}{n} x_2 - \cdots - \frac{1}{n} x_n,$$

$$\ldots$$

$$y_n = -\frac{1}{n} x_1 - \frac{1}{n} x_2 - \cdots + \frac{n-1}{n} x_n.$$

Chacun des résidus y_i est donc une combinaison linéaire des quantités $x_1, \ldots, x_n$ qui obéissent à la loi de Gauss. Il vérifie par conséquent lui-même cette loi avec le périmètre de précision k' tel que l'on a pour l'un quelconque des résidus

$$\frac{1}{k'^2} = \frac{1}{k^2} \frac{(n-1)^2 + 1^2 + \cdots + 1^2}{n^2} = \frac{n-1}{n} \frac{1}{k^2},$$

ce qui donne

$$k' = \sqrt{\frac{n}{n-1}}\, k.$$

La valeur quadratique moyenne de chacun des résidus est $\frac{1}{k'\sqrt{2}}$, or, la *loi empirique du hasard* nous conduit, si le nombre des observations est assez grand, à adopter comme valeur approchée de cette quantité *la racine carrée de la moyenne arithmétique des carrés des résidus effectifs*, soit $\sqrt{\frac{\Sigma y_i^2}{n}}$; dans ces conditions, une valeur

approchée de $\frac{1}{k\sqrt{2}}$ est $\sqrt{\frac{\Sigma y_i^2}{n-1}}$; l'erreur quadratique moyenne sur chacune des mesures est donc approximativement $\sqrt{\frac{\Sigma y_i^2}{n-1}}$, et, par conséquent, l'erreur quadratique moyenne sur leur moyenne arithmétique est

$$\mu = \sqrt{\frac{\Sigma y_i^2}{n(n-1)}}.$$

C'est une erreur qu'il y a une probabilité 0,6826 de ne pas dépasser en valeur absolue, il est raisonnable d'adopter ce chiffre comme mesure de l'*erreur à craindre*, nous avons cependant dit au § 27 pourquoi il serait illusoire d'espérer une amélioration indéfinie de la précision en multipliant outre mesure le nombre n des observations

42. Observations directes d'inégale précision. — Si nous avons des raisons d'attribuer aux observations qui ont donné les résultats z_1, z_2, . ., z_n des poids différents, proportionnels, ainsi que nous l'avons vu au § 27, aux carrés des modules de précision des observations, la valeur

$$z = \frac{p_1 z_1 + p_2 z_2 + \dots + p_n z_n}{p_1 + p_2 + \dots + p_n},$$

où p_1, p_2, . . , p_n désignent les poids, est celle qui assure le minimum de la somme des carrés des résidus multipliés par les carrés des modules de précision

$$k_1^2(Z - z_1)^2 + k_2^2(Z - z_2)^2 + \dots + k_n^2(Z - z_n)^2$$

Cette somme figure en exposant négatif dans l'expression de la probabilité *a posteriori* pour que la grandeur mesurée soit comprise entre Z et Z + dZ, moyennant l'hypothèse de Bayes, la valeur z est donc la plus probable, c'est celle que nous adopterons Du reste, un calcul tout semblable à celui du § 30 montre que z est en même temps la *valeur probable* de Z.

La détermination de l'*erreur à craindre* s'effectue comme au paragraphe précédent; les erreurs effectives $x_i = Z - z_i$ sont liées aux résidus $y_i = z - z_i$ et à l'erreur $x = Z - z$ commise en adoptant

la valeur z par les relations

$$y_1 = x_1 - z,\qquad y_2 = x_2 - z,\qquad \ldots,\qquad y_n = x_n - z,$$

et

$$p_1 y_1 + p_2 y_2 + \ldots + p_n y_n = 0.$$

Posons

$$p_1 + p_2 + \ldots + p_n = p,$$

et ajoutons membre à membre les $n+1$ relations ci-dessus après multiplication par $\frac{p-p_1}{p}$, $-\frac{p_2}{p}$, ..., $-\frac{p_n}{p}$, 1 respectivement.

Nous obtenons

$$y_1 = \frac{p-p_1}{p}x_1 - \frac{p_2}{p}x_2 - \ldots - \frac{p_n}{p}x_n$$

et, de même,

$$y_2 = -\frac{p_1}{p}x_1 + \frac{p-p_2}{p}x_2 - \ldots - \frac{p_n}{p}x_n,$$

$$\ldots\ldots\ldots\ldots$$

$$y_n = -\frac{p_1}{p}x_1 - \frac{p_2}{p}x_2 - \ldots + \frac{p-p_n}{p}x_n.$$

Il en résulte que chacun des résidus obéit à la loi de Gauss; soit k'_1 le paramètre correspondant pour y_1, nous avons

$$\frac{1}{k'^2_1} = \frac{(p-p_1)^2}{p^2}\frac{1}{k_1^2} + \frac{p_2^2}{p^2}\frac{1}{k_2^2} + \ldots + \frac{p_n^2}{p^2}\frac{1}{k_n^2}.$$

Désignons par ω le coefficient, inconnu, de proportionnalité des poids aux carrés des modules de précision ; si

$$\frac{p_1}{k_1^2} = \ldots = \frac{p_n}{k_n^2} = \omega$$

nous avons

$$\frac{1}{k'^2_1} = \frac{\omega}{p^2}\left[\frac{(p-p_1)^2}{p_1} + p_2 + \ldots + p_n\right] = \omega\left(\frac{1}{p_1} - \frac{1}{p}\right).$$

Les valeurs des carrés moyens des résidus sont donc

$$\frac{\omega}{2}\left(\frac{1}{p_1} - \frac{1}{p}\right),\quad \frac{\omega}{2}\left(\frac{1}{p_2} - \frac{1}{p}\right),\quad \ldots,\quad \frac{\omega}{2}\left(\frac{1}{p_n} - \frac{1}{p}\right),$$

et la valeur moyenne de l'expression

$$p_1 y_1^2 + \ldots + p_n y_n^2$$

est

$$\frac{\omega}{2}\left(n - \frac{p_1+p_2+\cdots+p_n}{p}\right) = (n-1)\frac{\omega}{2}.$$

Si nous adoptons la *valeur effective* de cette expression comme *valeur approchée de sa valeur moyenne*, nous avons pour l'inconnue auxiliaire ω la valeur approchée

$$\frac{2}{n-1}\sum p_i y_i^2.$$

et pour $\frac{1}{2h_1^2}$ la valeur approchée

$$\frac{\sum p_i y_i^2}{(n-1)p_1}$$

Ainsi, la valeur que nous sommes conduits à adopter pour moyenne quadratique de x_1 est

$$\mu_1 = \sqrt{\frac{p_1 y_1^2 + \cdots + p_n y_n^2}{(n-1)p_1}},$$

et la valeur quadratique moyenne de

$$z = \frac{p_1 x_1 + p_2 x_2 + \cdots + p_n x_n}{p_1 + p_2 + \cdots + p_n} = \frac{1}{p}\sum p_i x_i$$

est μ, défini par la relation

$$\mu^2 = \frac{p_1^2}{p^2}\mu_1^2 + \cdots + \frac{p_n^2}{p^2}\mu_n^2$$

Nous avons ainsi

$$\mu^2 = \frac{\sum p_i y_i^2}{n-1}\left[\frac{p_1+p_2+\cdots+p_n}{p^2}\right] = \frac{\sum p_i y_i^2}{(n-1)p},$$

et l'*erreur à craindre* en adoptant le résultat z défini plus haut est

$$\mu = \sqrt{\frac{\sum p_i y_i^2}{(n-1)p}}.$$

43. Examen du même problème au point de vue de la statistique mathématique. — Si l'on a appliqué la méthode du § 41 à un nombre d'observations assez élevé, il est possible de se rendre compte de la mesure dans laquelle la loi de Gauss, que l'on a supposée vérifiée par les erreurs des observations, est vérifiée effectivement Ce problème de l'*ajustement* d'une série de résultats à la loi de Gauss

peut être envisagé pour toutes sortes de *séries statistiques* et rentre dans le cadre de la *statistique mathématique* (*voir* le fascicule IV du Tome I du présent Traité).

Il arrive que les nombres en présence desquels on se trouve ne vérifient pas la loi de Gauss d'une manière satisfaisante; on est ainsi conduit à chercher des formes plus générales de lois de probabilité, vis-à-vis desquelles la loi de Gauss constitue seulement une première approximation comme, dans la Physique des gaz, la loi des gaz parfaits vis-à-vis d'équations d'état plus compliquées et mieux adaptées au cas des gaz réels.

A cet égard, un grand parti peut être tiré du développement de la fonction des probabilités élémentaires, multipliée par un facteur exponentiel convenable, en série de polynomes d'Hermite-Tchebychef. Cette série de polynomes est souvent appelée la série de *Gram-Charlier*. On peut établir sa convergence et sa validité pour toute fonction vérifiant les conditions de Dirichlet ([1]). Nous nous bornerons à quelques indications très sommaires, renvoyant le lecteur, pour plus de développements, aux fascicules III et IV (Tome I du présent Traité).

Soit $f(x)$ la fonction des probabilités élémentaires relative à la variable éventuelle x, qui, dans le cas actuel, est une erreur d'observation. Effectuons le changement de variable $x = a + \frac{u}{h}$, a et h étant deux constantes que nous déterminerons plus loin de manière à simplifier les calculs. Si nous admettons la validité du développement de la fonction

$$g(u) = \frac{\sqrt{\pi}}{h} f\left(a + \frac{u}{h}\right) e^{u^2}$$

en série de polynomes d'Hermite

$$c_0 P_0(u) + \frac{c_1}{1} P_1(u) + \frac{c_2}{2!} P_2(u) + \dots,$$

avec

$$P_n(u) = (-1)^n e^{u^2} \frac{d^n}{du^n} (e^{-u^2}),$$

([1]) *Voir* GALBRUN, *Bulletin de la Société mathématique*, 1913. Le premier Mémoire sur la question est de GRAM (*Journal de Crelle*, 1883).

nous écrivons l'identité

$$f\left(a-\frac{u}{h}\right)=\frac{h}{\sqrt{\pi}}e^{-u^2}\left[c_0 P_0(u)-\frac{c_1}{1}P_1(u)+\frac{c_2}{2!}P_2(u)-\ldots\right]$$

où $P_n(u)$ est le produit par 2^n du polynome Q_n envisagé au § 35.

Les coefficients c_n se calculent d'une manière analogue à celle qui donne les coefficients du développement d'une fonction en série de Fourier.

Il résulte, en effet, des indications du § 28, qu'avec les notations actuelles, l'intégrale

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}P_m(u)P_n(u)e^{-u^2}du$$

a pour valeur zéro pour $m \neq n$, et $2^n n!$ pour $m = n$

On a ainsi immédiatement

$$2^n c_n=\frac{1}{h}\int_{-\infty}^{+\infty}f\left(a-\frac{u}{h}\right)P_n(u)\,du$$

Dans ces conditions, choisissons a et h de manière que, dans le développement ci-dessus, les coefficients c_1 et c_2 soient tous les deux nuls. Comme

$$P_0(u)=1,\qquad P_1(u)=2u,\qquad P_2(u)=(2u)^2-2,$$

nous avons, avec la variable x telle que $u=h(x-a)$,

$$\int_{-\infty}^{+\infty}(x-a)f(x)\,dx=0$$

et

$$\int_{-\infty}^{+\infty}[4h^2(x-a)^2-2]f(x)\,dx=0$$

L'intégrale

$$\int_{-\infty}^{+\infty}f(x)\,dx$$

ayant pour valeur 1 de par la nature même de la fonction $f(x)$, nous voyons que

$$a=\int_{-\infty}^{+\infty}x f(x)\,dx.$$

c'est-à-dire que a est la *valeur probable de* x ; cette valeur est nulle s'il n'y a pas d'erreur systématique

Introduisons les moments

$$m_h = \int_{-\infty}^{+\infty} (x - a)^h f(x)\, dx$$

calculés « par rapport » à cette valeur probable, nous avons

$$h = \frac{1}{\sqrt{2 m_2}},$$

et le changement de variable

$$x = a + \frac{u}{h},$$

qu'on peut écrire

$$u = \frac{x - \mu_1}{\sqrt{2 m_2}},$$

revient, à un facteur $\frac{1}{\sqrt{2}}$ près, à la *réduction* de la loi de probabilité telle que nous l'avons utilisée au Chapitre V.

Le développement se réduit ainsi à la forme

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 u^2} \left[1 + \frac{c_3}{3!} P_3(u) + \ldots \right],$$

le coefficient c_0 étant égal à l'unité Dans le cas de la loi de Gauss, tous les coefficients à partir de c_3 sont nuls, dans des cas plus généraux, on aura une représentation de $f(x)$ en limitant la série entre crochets à un certain nombre de ses premiers termes.

Le calcul des coefficients est lié à la considération des moments m_h, en effet, on établit par récurrence l'expression

$$P_h(u) = (2u)^h - \frac{h(h-1)}{1}(2u)^{h-2} + \frac{h(h-1)(h-2)(h-3)}{1 \cdot 2}(2u)^{h-4} - \ldots,$$

il en résulte, d'après la détermination donnée plus haut des coefficients c, que

$$\begin{aligned} 2^h c_h &= \int_{-\infty}^{+\infty} f(x)\, P_h[h(x-a)]\, dx \\ &= (2h)^h m_h - \frac{h(h-1)}{1}(2h)^{h-2} m_{h-2} + \ldots, \end{aligned}$$

on peut donc aisément faire l'ajustement approché d'une série de mesures au développement de Gram-Charlier convenablement limité.

II — COMBINAISON D'OBSERVATIONS INDIRECTES OU CONDITIONNELLES.

44. Observations indirectes. — Le problème de la combinaison des observations indirectes a été énoncé au § 19. il consiste à déduire des mesures des quantités

$$U_1 = f_1(X, Y, Z, \ldots)$$
$$U_2 = f_2(X, Y, Z, \ldots)$$
$$\ldots$$
$$U_n = f_n(X, Y, Z, \ldots)$$

les valeurs les plus plausibles des inconnues X, Y, Z, ..., dont le nombre k est inférieur à n.

Soient $L_1, L_2, \ldots, L_n$ les résultats des mesures, si nous admettons d'une part, comme au § 19, l'hypothèse de Bayes d'après laquelle, dans l'ignorance où nous sommes de la loi de probabilité *a priori* des valeurs X, Y, Z, ..., le plus simple est de supposer la fonction φ constante, et si nous supposons, d'autre part que les mesures directes obéissent à la loi de Gauss avec les paramètres de précision $h_1, h_2, \ldots, h_n$, la probabilité pour que les inconnues aient des valeurs respectivement comprises entre X et $X + dX$, Y et $Y + dY$, Z et $Z + dZ$, ... est proportionnelle à $e^{-\Omega}$, où

$$\Omega = h_1^2(f_1 - L_1)^2 + h_2^2(f_2 - L_2)^2 + \ldots + h_n^2(f_n - L_n)^2.$$

Interprétant comme plus haut l'énoncé du problème dans le sens qui consiste à adopter les valeurs *les plus probables* des inconnues, nous sommes donc conduits à chercher le minimum de Ω; Ω est la somme des carrés des résidus, carrés multipliés respectivement par les carrés des modules de précision, c'est-à-dire par les poids des observations.

C'est là l'énoncé général de ce qu'on appelle la *méthode des moindres carrés,* la méthode est ainsi justifiée d'une manière directe en admettant la loi de Gauss, l'hypothèse de Bayes, et en choisissant comme valeurs les plus plausibles des inconnues les valeurs *les plus probables*.

Ces valeurs sont aussi les *valeurs probables*, ainsi que l'a montré

Poincaré, pourvu que les relations entre les U_i et les inconnues X, Y, Z, ... soient *linéaires*, nous verrons dans la troisième Partie du présent fascicule que cette hypothèse ne restreint la généralité qu'en apparence, pourvu que les erreurs soient très petites.

En effet, soient alors $x, y, z, \ldots$ les valeurs adoptées qui rendent Ω minimum, si les relations $V_i = f_i(X, Y, Z, \ldots)$ sont linéaires, Ω est du second degré en X, Y, Z, ..., et les valeurs $x, y, z, \ldots$ sont les coordonnées du centre de l'hypersurface $\Omega = 0$. Il en résulte que Ω est la somme d'une forme quadratique homogène en $X - x$, $Y - y$, $Z - z, \ldots$, et d'une fonction $\Omega(x, y, z, \ldots)$ indépendante de X, Y, Z, ..., représentant le minimum Ω_0 de Ω; écrivons

$$\Omega = \Phi + \Omega_0.$$

La valeur probable de X, par exemple, est

$$\overline{X} = \frac{\underbrace{\int \cdots \int}_{k} X e^{-\Omega} dX\, dY\, dZ \ldots}{\underbrace{\int \cdots \int}_{k} e^{-\Omega} dX\, dY\, dZ \ldots};$$

pour démontrer que $\overline{X} = x$, il suffit d'établir que

$$\underbrace{\int \cdots \int}_{k} (X - x) e^{-\Omega} dX\, dY\, dZ \ldots = 0.$$

Or, ce résultat est évident, car l'intégrale est étendue à tout l'espace, et la fonction à intégrer est impaire par rapport à $X - x$, paire par rapport à $Y - y$, $Z - z, \ldots$.

45. **Interpolation parabolique.** — Soit une fonction inconnue $f(x)$, d'origine expérimentale, dont on connaît seulement les valeurs u_1, $u_2, \ldots, u_n$, mesurées dans des observations portant sur $f(x)$ pour $x = x_1, x_2, \ldots, x_n$. On connaît ainsi n points de la courbe $y = f(x)$.

Le problème de la représentation analytique, aussi approchée que possible, du phénomène étudié est le *problème général de l'interpolation;* il présente évidemment une large part d'arbitraire, correspondant aux diverses formes de fonctions $f(x)$ susceptibles d'être

envisagées. Son énoncé devient plus précis dans le cas, pratiquement très important, où l'on recherche la plus simple possible des représentations par un polynome entier

$$f(x) = a_0 + a_1 x + \dots + a_q x^q$$

on a alors affaire à l'*interpolation parabolique*.

Le degré q du polynome $f(x)$ est inférieur à $n - 1$; s'il était égal à $n - 1$, le polynome serait parfaitement déterminé, et la formule d'interpolation de Lagrange donnant immédiatement son expression est bien connue; en fait, on essaie de trouver les coefficients a_0, a_1, ..., a_q d'un polynome de degré très inférieur au nombre des observations et représentant le mieux possible le phénomène étudié.

Les observations étant entachées d'erreurs, les différences

$$f(x_i) - u_i$$

ne sont pas nulles, même si la fonction $f(x)$ est la représentation rigoureusement exacte du phénomène. Nous pouvons considérer le problème comme portant sur les observations indirectes des quantités a_0, a_1, ..., a_q par l'intermédiaire de u_1, u_2, ..., u_n, et si nous admettons que les observations sont d'égale précision, nous adopterons les valeurs des coefficients assurant le minimum de la somme des carrés des résidus

$$\Omega = \sum_i (a_0 + a_1 x_i + \dots + a_q x_i^q - u_i)^2$$

Le problème aboutit au système des équations linéaires

$$\frac{\partial \Omega}{\partial a_0} = 0, \qquad \frac{\partial \Omega}{\partial a_1} = 0, \qquad \dots, \qquad \frac{\partial \Omega}{\partial a_q} = 0$$

qui s'écrivent

$$\begin{aligned} \Sigma u_i &= \Sigma f(x_i), \\ \Sigma u_i x_i &= \Sigma x_i f(x_i), \\ &\dots\dots \\ \Sigma u_i x_i^q &= \Sigma x_i^q f(x_i), \end{aligned}$$

et expriment, par conséquent, que les *sommes des moments* d'ordres 0, 1, 2, ..., q des masses u_1, u_2, ..., u_n, placées aux points x_1, x_2, ..., x_n, sont les mêmes pour ces masses observées ou

pour ces masses calculées au moyen du polynome $f(x)$; d'où le nom de *methode des moments* donné souvent à ce procédé de calcul

Si, une fois les coefficients calculés, les valeurs des résidus sont trop fortes vis-à-vis de l'approximation désirée, on doit reprendre les recherches *avec une valeur plus grande de l'exposant q.*

La méthode suivante, indiquée par Poincaré, permet, en ce cas, *d'utiliser la totalité des calculs antérieurs.*

Posons

$$F(x) = (x - x_1)(x - x_2) \ldots (x - x_n),$$

et mettons la fraction rationnelle

$$\frac{F'(x)}{F(x)}$$

sous la forme du développement en fraction continue

$$\frac{F'}{F} = \cfrac{1}{Q_1 + \cfrac{1}{Q_2 + \cdots + \cfrac{1}{Q_n}}}.$$

On appelle réduite de rang h la fraction

$$\frac{N_h}{D_h} = \cfrac{1}{Q_1 + \cfrac{1}{Q_2 + \cdots + \cfrac{1}{Q_h}}},$$

et l'on sait que si l'on effectue les divisions successives

$$\begin{aligned} F &= Q_1 F' + R_1, \\ F' &= Q_2 R_1 + R_2, \\ &\ldots \\ R_{h-2} &= Q_h R_{h-1} + R_h, \\ &\ldots \end{aligned}$$

on démontre par récurrence l'identité

$$(h) \qquad N_h F - D_h F' = R_h.$$

Les dénominateurs D_h possèdent cette propriété que, pour h donné et μ inférieur à h, on a

$$x_1^\mu D_h(x_1) + x_2^\mu D_h(x_2) + \ldots + x_n^\mu D_h(x_n) = 0.$$

En effet, la fraction rationnelle $x^{\mu}\frac{D_h F'}{F}$ s'écrit, en vertu de l'identité (h),

$$N_h x^{\mu} - \frac{x^{\mu} R_h}{F},$$

et se décompose en eléments simples sous la forme

$$N_h x^{\mu} - \frac{A_1}{x - x_1} - \frac{A_2}{x - x_2} - \cdots - \frac{A_n}{x - x_n}$$

Si l'on chasse les dénominateurs dans l'identite

$$\frac{x^{\mu} R_h}{F} = \sum \frac{A_i}{x - x_i},$$

on a, au premier membre, un polynome de degré $\mu + n - 1 - h$, inférieur à $n - 1$ si μ est inférieur à h, au second membre, un polynome de degré $n - 1$ dont le terme en x^{n-1} a pour coefficient ΣA_i donc

$$\Sigma A_i = 0$$

résidu relatif au pôle x_i pour la fraction

$$x^{\mu} \frac{D_h F'}{F} = \frac{P}{Q},$$

sa valeur est

$$\frac{P(x_i)}{Q'(x_i)} = x_i^{\mu} D_h(x_i),$$

la relation $\Sigma A_i = 0$ s'ecrit bien

$$(\mu) \qquad x_1^{\mu} D_h(x_1) + x_2^{\mu} D_h(x_2) + \cdots + x_n^{\mu} D_h(x_n) = 0$$

Il s'ensuit immédiatement que, si $\theta(x)$ est un polynome quelconque de degré inférieur à h, on a

$$\Sigma\, \theta(x_i) D_h(x_i) = 0,$$

et que deux dénominateurs D_h, D_k possèdent la propriété d'orthogonalité

$$\Sigma\, D_h(x_i) D_k(x_i) = 0$$

Dans ces conditions, mettons le polynome $f(x)$ sous la forme

$$\lambda_0 D_0 + \lambda_1 D_1(x) + \ldots + \lambda_q D_q(x), \qquad \text{où } D_0 = 1:$$

la somme des carrés des résidus se réduit, d'après la relation (ν) pour $\mu = 0$ et la propriété d'orthogonalité ci-dessus, à

$$\Omega = \sum_1^n u_i^2 + \sum_0^q \lambda_h^2 \sum_1^n D_h^2(x_i) - 2\sum_0^q \lambda_h \sum_1^n u_i D_h(x_i)$$

La dérivée par rapport à λ_h s'écrit

$$2\lambda_h \sum_1^n D_h^2(x_i) - 2\sum_1^n u_i D_h(x_i),$$

donc les valeurs de $\lambda_0, \lambda_1, \ldots, \lambda_q$ sont données par les formules

$$\lambda_h = \frac{u_1 D_h(x_1) + \ldots + u_n D_h(x_n)}{D_h^2(x_1) + \ldots + D_h^2(x_n)}.$$

Si l'on constate des valeurs trop fortes pour les résidus, il suffit d'ajouter à la somme déjà calculée

$$\lambda_0 D_0 + \ldots + \lambda_q D_q(x)$$

un ou plusieurs termes supplémentaires sans avoir à recommencer tous les calculs.

Nous nous bornerons à ces indications, le problème mathématique général de l'approximation d'une fonction par un polynome de degré donné sortant du cadre de cet Ouvrage ([1])

46. Observations conditionnelles. — Reprenons, pour terminer ce Chapitre, le problème de compensation d'observations conditionnelles énoncé au § 20. Les equations de condition

$$g_1(U_1, U_2, \ldots, U_n) = 0,$$
$$g_2(U_1, U_2, \ldots, U_n) = 0,$$
$$\ldots\ldots\ldots,$$
$$g_p(U_1, U_2, \ldots, U_n) = 0,$$

laissent le point $(U_1, U_2, \ldots, U_p)$ arbitraire dans une multiplicité à $n - p$ dimensions; on peut exprimer, par conséquent, $U_1, U_2, \ldots, U_n$

([1]) *Voir* Émile Borel, *Leçons sur les fonctions de variables réelles.*

en fonction de $q = n - p$ paramètres, qui sont tout naturellement les k inconnues X, Y, Z, ... dans le cas du § 19 où les inconnues sont indépendantes et qui peuvent, dans tous les cas, jouer le rôle d'inconnues auxiliaires possédant cette indépendance

Soient $X_1, X_2, \ldots, X_q$ de tels paramètres. les divers systèmes de valeurs des grandeurs mesurées $U_1, U_2, \ldots, U_n$, compatibles avec les équations de condition, correspondent aux divers points d'une multiplicité à q dimensions $X_1, X_2, \ldots, X_q$. La probabilité élémentaire relative à un système de valeurs acceptables des U_i sera représentée par une expression portant sur les valeurs des paramètres indépendants $X_1, X_2, \ldots, X_q$. Dès lors, en reprenant exactement l'étude du § 19, on voit que la probabilité d'un système de valeurs acceptables des U_i a la forme

$$\frac{\psi\theta_1\theta_2\ldots\theta_n\,dX_1\,dX_2\ldots dX_q}{\underbrace{\int\ldots\int}_{q}\psi\theta_1\theta_2\ldots\theta_n\,dX_1\,dX_2\ldots dX_q},$$

où $\psi(X_1, X_2, \ldots, X_q)\,dX_1\,dX_2\ldots dX_q$ désigne la probabilité élémentaire *a priori* relative aux valeurs $X_1, X_2, \ldots, X_q$, et où $\theta_1(L_1, U_1), \theta_2(L_2, U_2), \ldots, \theta_n(L_n, U_n)$ sont les fonctions des probabilités élémentaires caractérisant les lois respectives d'erreur des observations.

Admettant une fois de plus l'hypothèse de Bayes, d'après laquelle $\psi = \text{const.}$, et supposant que les mesures directes satisfont à la loi de Gauss avec les paramètres de précision $k_1, k_2, \ldots, k_n$, nous sommes conduits, comme au § 19, à une probabilité élémentaire *a posteriori* des valeurs $X_1, X_2, \ldots, X_q$ proportionnelle à $e^{-\Omega}$, avec

$$\Omega = k_1(U_1 - L_1)^2 + k_2(U_2 - L_2)^2 + \ldots + k_n(U_n - L_n)^2$$

Les valeurs les plus probables de $U_1, U_2, \ldots, U_n$, compte tenu des équations de condition, s'obtiennent donc par la méthode donnant les valeurs de n variables qui assurent le minimum d'une fonction de ces variables lorsqu'il existe entre elles un certain nombre de relations.

D'après cette méthode, la relation

$$k_1^2(U_1 - L_1)\,dU_1 + k_2^2(U_2 - L_2)\,dU_2 + \ldots + k_n^2(U_n - L_n)\,dU_n$$

est la conséquence, pour le voisinage des valeurs U_1, U_2, .., U_n cherchées, des equations de condition et des relations

$$\frac{\partial g_1}{\partial U_1} dU_1 + \quad + \frac{\partial g_1}{\partial U_n} dU_n = 0,$$

$$\frac{\partial g_2}{\partial U_1} dU_1 + \quad + \frac{\partial g_2}{\partial U_n} dU_n = 0,$$

$$\ldots\ldots\ldots\ldots,$$

$$\frac{\partial g_p}{\partial U_1} dU_1 + \quad + \frac{\partial g_p}{\partial U_n} dU_n = 0$$

TROISIÈME PARTIE.

MÉTHODE DES MOINDRES CARRES

CHAPITRE IX.

COMBINAISON D'OBSERVATIONS DIRECTES.

47. **Principe du moindre risque d'erreur.** — On peut se faire une idée de la précision d'une mesure pour laquelle la loi de probabilité des erreurs est caractérisée par la fonction des probabilités totales $F(x)$ [ou éventuellement, la fonction des probabilités élémentaires $f(x)$], par la valeur de l'intégrale

$$J = \int g(x)\,dF(x),$$

où $g(x)$ est une certaine fonction de l'erreur. Pour que la valeur numérique de J, espérance mathématique d'un joueur qui paierait $g(x)$ pour une erreur x, caractérise, malgré la grande part restante d'arbitraire, la précision des mesures, il est indispensable que la fonction $g(x)$ soit *paire* et *croissante* avec $|x|$ On aboutit ainsi à une définition très large de ce qu'on peut appeler le *risque d'erreur*.

Si la loi de probabilité des erreurs est la loi de Gauss, il est facile de voir que, quelle que soit la fonction $g(x)$ vérifiant les conditions ci-dessus, le risque d'erreur est d'autant plus grand que le paramètre de précision est plus petit.

On a, en effet,

$$J = \frac{k}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-k^2x^2} g(x)\,dx$$

ou, en posant $hx = u$, et tenant compte de la symétrie,

$$J = \frac{2}{\sqrt{\pi}} \int_0^{+\infty} e^{-u^2} g\left(\frac{u}{h}\right) du,$$

la fonction $g(x)$ étant croissante, J est d'autant plus grand que h est plus petit.

La méthode des moindres carrés, conséquence logique de l'hypothèse de Bayes, de la loi de Gauss et de l'interprétation consistant à adopter comme valeurs les plus plausibles les valeurs les plus probables, peut être justifiée, dans les divers cas où nous l'avons envisagée, par la méthode basée sur le *principe du moindre risque d'erreur*, qui consiste à chercher les valeurs des inconnues assurant le minimum du risque J correspondant, en adoptant pour mesurer ce risque *la valeur moyenne du carré de l'erreur*.

Gauss, qui énonce ce principe dans sa *Theoria combinationis* (1), appelée souvent « deuxième théorie de Gauss », déclare qu'en ce qui concerne la mesure du risque d'erreur, parmi le nombre infini de fonctions qui remplissent les conditions requises, « il semble naturel de choisir la plus simple, qui est, sans contredit, le carré de l'erreur ».

Nous allons reprendre l'étude des divers cas, énoncés au Chapitre III et étudiés déjà au Chapitre VIII, du problème de la combinaison des observations, en nous plaçant au point de vue de ce principe du minimum de l'erreur quadratique moyenne et afin d'examiner plus en détail la technique des opérations. Les hypothèses que nous ferons constamment sont celles de *l'exclusion des erreurs systématiques* et de *la petitesse des erreurs accidentelles*. Elles entraînent, d'après la théorie des erreurs partielles, que les erreurs vérifient très approximativement la loi de Gauss, donc, conformément à la remarque finale du § 31, notre exposé ne sera qu'en apparence indépendant de cette loi; mais il nous paraît intéressant de traiter, en tout état de cause, la question selon ce point de vue, qui présente l'avantage d'affranchir la méthode des moindres carrés de la théorie des probabilités des causes et de l'hypothèse de Bayes.

---

(1) *Voir* la traduction française par J. BERTRAND, sous le titre : *Méthode des moindres carrés* (Paris, Gauthier-Villars).

48. Observations directes d'inégale précision — Supposons que les mesures ayant donné pour Z les valeurs $z_1, z_2, \ldots, z_n$ ont des poids $p_1, p_2, \ldots, p_n$, connus à un facteur constant près et respectivement proportionnels aux inverses des carrés moyens $m_1^2, m_2^2, \ldots, m_n^2$ des erreurs commises. Il est naturel de chercher la valeur de moindre erreur quadratique moyenne parmi celles dont l'expression générale est

$$z = \lambda_1 z_1 + \lambda_2 z_2 + \ldots + \lambda_n z_n,$$

où les coefficients λ_i ont pour somme l'unité. En effet, si l'on multiplie les z_i par une constante, z doit être multiplié par cette constante; si l'on ajoute aux z_i une même constante, z doit être accru de cette constante, enfin, si tous les z_i sont égaux, z doit être égal à leur valeur commune

Cherchons, parmi toutes les valeurs approchees ainsi obtenues, celle dont l'erreur quadratique est la plus faible. Soient

$$x_1 = Z - z_1$$
$$x_2 = Z - z_2,$$
$$\ldots\ldots$$
$$x_n = Z - z_n$$

les erreurs commises dans les observations l'erreur sur

$$z = \lambda_1 z_1 + \lambda_2 z_2 + \ldots + \lambda_n z_n$$

a pour expression

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n,$$

son carré est

$$\Sigma \lambda_i^2 x_i^2 + 2 \Sigma \lambda_i \lambda_j x_i x_j$$

Cherchons la valeur moyenne de ce carré.

Celle de chacun des produits $x_i x_j$ est nulle, d'après l'hypothese de l'exclusion de toute erreur systématique; donc, la valeur probable de x^2 est

$$\lambda_1^2 m_1^2 + \lambda_2^2 m_2^2 + \ldots + \lambda_n^2 m_n^2$$

Les valeurs de λ_i rendant cette somme minimum doivent vérifier les deux conditions

$$\Sigma\, d\lambda_i = 0,$$

$$\Sigma\, m_i^2 \lambda_i\, d\lambda_i = 0,$$

qui sont donc nécessairement équivalentes. Cela exige

$$m_1^2\lambda_1 = m_2^2\lambda_2 = \dots = m_n^2\lambda_n,$$

et entraîne

$$\lambda_i = \frac{\frac{1}{m_i^2}}{\Sigma\frac{1}{m_i^2}},$$

de sorte que la valeur à adopter est

$$z = \frac{\frac{z_1}{m_1^2} + \frac{z_2}{m_2^2} + \dots + \frac{z_n}{m_n^2}}{\frac{1}{m_1^2} + \frac{1}{m_2^2} + \dots + \frac{1}{m_n^2}};$$

c'est celle dont il est question aux §§ 27 et 41, dans le cas où la loi de Gauss est admise, on peut l'écrire $\frac{\Sigma p_i z_i}{\Sigma p_i}$, et c'est celle qui réalise le minimum de la somme $2\Omega = \Sigma p_i(z - z_i)^2$. *La méthode des moindres carrés* se trouve ainsi justifiée.

49. Calcul de l'erreur à craindre. — La valeur probable de x^2 admet, dans ces conditions, un minimum égal à $\frac{1}{\Sigma\frac{1}{m_i^2}}$.

Examinons comment on peut se faire une idée de la valeur de ce minimum lorsque l'on connaît, non pas les quantités $\frac{1}{m_i^2}$, mais des poids $p_1, p_2, \dots, p_n$ qui leur sont proportionnels.

Les résidus $y_1 = z - z_1, y_2 = z - z_2, \dots, y_n = z - z_n$ sont liés aux erreurs véritables $x_1, x_2, \dots, x_n$ par les relations, déjà écrites au § 42, d'où il résulte, en posant $\Sigma p_i = p$, que

$$y_1 = \frac{p - p_1}{p}x_1 - \frac{p_2}{p}x_2 - \dots - \frac{p_n}{p}x_n.$$

La valeur moyenne de y_1^2 est donc

$$m'^2_1 = \frac{1}{p^2}[(p - p_1)^2 m_1^2 + p_2^2 m_2^2 + \dots + p_n^2 m_n^2],$$

ou, en posant

$$p_1 m_1^2 = p_2 m_2^2 = \dots = p_n m_n^2 = \lambda,$$

quantité souvent appelée carré moyen de l'erreur d'une observation

de poids *un*,

$$m_1'^2 = \frac{\lambda}{p}\left[\frac{(p-p_1)}{p_1} + p_2 + \cdots + p_n\right] = \lambda\left(\frac{1}{p_1} - \frac{1}{p}\right)$$

La valeur moyenne de

$$X = p_1 y_1'^2 + \cdots + p_n y_n'^2$$

est, dans ces conditions,

$$p_1 m_1'^2 + \cdots + p_n m_n'^2 = \lambda\left(n - \frac{\Sigma p_i}{p}\right) = (n-1)\lambda$$

et nous avons, d'après la loi empirique du hasard, une valeur approchée de λ en remplaçant cette valeur moyenne par la valeur constatée.

Adoptant ainsi

$$\lambda = \frac{p_1 y_1^2 + \cdots + p_n y_n^2}{n-1},$$

nous avons

$$\frac{1}{m_1^2} = \frac{p_1}{\lambda} = \frac{(n-1)p_1}{\Sigma p_i y_i^2}$$

et

$$\Sigma \frac{1}{m_i^2} = \frac{(n-1)p}{\Sigma p_i y_i^2}.$$

Il s'ensuit que la valeur moyenne de x^2 est $\frac{\Sigma p_i y_i^2}{(n-1)p}$, et que l'erreur quadratique moyenne sur $z = \frac{\Sigma p_i z_i}{\Sigma p_i}$ est donnée par l'expression

$$\mu = \sqrt{\frac{\Sigma p_i y_i^2}{(n-1)p}},$$

déjà obtenue au § 42 en partant de la loi de Gauss.

Si les observations sont d'égale précision, z se réduit à la moyenne arithmétique

$$\frac{z_1 + z_2 + \cdots + z_n}{n},$$

et μ a l'expression

$$\sqrt{\frac{\Sigma y_i^2}{n(n-1)}}$$

obtenue au § 41.

CHAPITRE X.

COMBINAISON D'OBSERVATIONS INDIRECTES

50 **Équations d'erreur linéaires.** — Reprenons le problème de la meilleure détermination des inconnues X, Y, Z, ... par les mesures des fonctions $U_1(X, Y, Z, \ldots)$, $U_2(X, Y, Z, \ldots)$, ... $U_n(X, Y, Z, \ldots)$, dont les résultats sont L_1, L_2, ..., L_n. Faisons sur les observations considérées les hypothèses déjà énoncées au § 47, d'après lesquelles : 1° il n'y a pas d'erreurs systématiques ; 2° les erreurs accidentelles sont très petites.

Dans ces conditions, le domaine de variation possible des quantités X, Y, Z, ... est très restreint ; si x_0, y_0, z_0, ... sont les coordonnées d'un point de ce domaine, les erreurs commises en adoptant ces coordonnées comme valeurs approchées des inconnues sont très petites ; nous préciserons cette hypothèse en admettant que les carrés de ces erreurs et leurs produits deux à deux sont négligeables devant les erreurs elles-mêmes.

Posons donc

$$X = x_0 + x, \qquad Y = y_0 + y, \qquad Z = z_0 + z, \ldots$$

et supposons les fonctions $f_1, f_2, \ldots, f_n$ développables suivant les puissances de $x, y, z, \ldots$ D'après ce qui précède, les développements pourront se réduire à leurs termes de degrés 0 et 1 et s'écrire

$$u_1 + a_1 x + b_1 y + c_1 z + \ldots,$$
$$\ldots\ldots\ldots\ldots,$$
$$u_n + a_n x + b_n y + c_n z + \ldots.$$

Moyennant cette importante simplification, on peut remplacer les

équations $f_1 = L_1, f_2 = L_2, \dots, f_n = L_n$ par les suivantes

$$(1)\quad \begin{cases} a_1 x + b_1 y + c_1 z + \dots - l_1 = 0 \\ a_2 x + b_2 y + c_2 z + \dots - l_2 = 0 \\ \dots \\ a_n x + b_n y + c_n z + \dots - l_n = 0 \end{cases}$$

où l_i désigne la différence $L_i - f_i(x_0, y_0, z_0, \dots)$.

Ces équations ne sont pas compatibles ; si nous appelons v_1, v_2, ..., v_n les formes linéaires respectivement égalées ainsi à l_1, l_2, ..., l_n, nous serons conduits à adopter des valeurs des inconnues $x, y, z, \dots$ qui donneront aux écarts $\varepsilon_1 = v_1 - l_1$, $\varepsilon_2 = v_2 - l_2$, ..., $\varepsilon_n = v_n - l_n$ des valeurs généralement non nulles appelées *résidus*.

Les équations

$$(L)\quad \begin{cases} a_1 x + b_1 y + c_1 z + \dots - l_1 = \varepsilon_1 \\ a_2 x + b_2 y + c_2 z + \dots - l_2 = \varepsilon_2 \\ \dots \\ a_n x + b_n y + c_n z + \dots - l_n = \varepsilon_n \end{cases}$$

dans lesquelles $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ désignent ces résidus, sont les *équations d'erreur*; on peut les considérer comme formant, par rapport aux inconnues $x, y, z, \dots$ un système d'*équations linéaires compatibles*. Dans l'étude qui va suivre, nous ferons constamment usage des notations, introduites par Gauss, consistant à désigner par $[aa]$, $[bb]$, ..., les sommes de carrés

$$a_1^2 + a_2^2 + \dots + a_n^2, \quad b_1^2 + b_2^2 + \dots + b_n^2,$$

et par $[ab]$, $[al]$, ..., les sommes de termes rectangles telles que

$$a_1 b_1 + a_2 b_2 + \dots + a_n b_n, \quad a_1 l_1 + a_2 l_2 + \dots + a_n l_n.$$

51. Détermination de x par le minimum de l'erreur quadratique moyenne. — Examinons d'abord le cas où les observations directes, dont nous pouvons considérer que les résultats sont l_1, l_2, ..., l_n, sont d'*égale précision*, et désignons par m la valeur commune des erreurs quadratiques moyennes correspondantes.

Portons notre attention sur l'inconnue x; il existe, puisque n est supérieur à k, une infinité de systèmes de multiplicateurs $\rho_1, \rho_2, \dots, \rho_n$ tels que l'on ait identiquement

$$(2)\quad x = \rho_1 v_1 + \rho_2 v_2 + \dots + \rho_n v_n.$$

Si l'on ajoute les équations (1) membre à membre, après multiplication par $\rho_1, \rho_2, \ldots, \rho_n$, on obtient

$$x = \rho_1 l_1 + \rho_2 l_2 + \quad + \rho_n l_n$$

Conformément au principe adopté, nous cherchons celles des quantités $[\rho l]$ ainsi obtenues, les ρ_i vérifiant l'identité (2), *pour laquelle l'erreur quadratique moyenne est minimum.*

Cette erreur quadratique moyenne, en vertu de l'hypothèse d'exclusion préalable des erreurs systématiques, a pour carré

$$m^2(\rho_1^2 + \rho_2^2 + \quad + \rho_n^2),$$

les multiplicateurs doivent donc réaliser le minimum de la somme

$$\omega = \rho_1^2 + \rho_2^2 + \quad + \rho_n^2 = [\rho\rho],$$

tout en vérifiant les relations

$$[a\rho] = a_1\rho_1 + a_2\rho_2 + \quad + a_n\rho_n = 1, \qquad [b\rho] = 0, \qquad [c\rho] = 0,$$

par lesquelles se traduit l'identité (2).

Dans le voisinage des valeurs cherchées des ρ_i, on devra donc avoir

$$\rho_1 d\rho_1 + \rho_2 d\rho_2 + \quad + \rho_n d\rho_n = 0,$$

pourvu que

$$a_1 d\rho_1 + a_2 d\rho_2 + \quad + a_n d\rho_n = 0,$$
$$b_1 d\rho_1 + b_2 d\rho_2 + \quad + b_n d\rho_n = 0,$$

Il en résulte que chaque ρ_i est une combinaison linéaire de la forme

$$\rho_i = \alpha a_i + \beta b_i + \gamma c_i + \quad ,$$

$\alpha, \beta, \gamma, \ldots$ désignant k inconnues auxiliaires

Portant ces valeurs dans les relations exprimant l'identité (2), on obtient le système

$$(S) \quad \begin{cases} [aa]\alpha + [ab]\beta + [ac]\gamma + \ldots = 1, \\ [ba]\alpha + [bb]\beta + [bc]\gamma + \ldots = 0, \\ [ca]\alpha + [cb]\beta + [cc]\gamma + \ldots = 0, \end{cases}$$

de k équations linéaires en $\alpha, \beta, \gamma, \ldots$, qui fournit la solution du problème.

Supposons le déterminant Δ des coefficients des inconnues différent de zéro et désignons par (aa), (ab), ..., les *mineurs normalisés*, quotients par Δ des coefficients de $[aa]$, $[ab]$, ..., la résolution du système (S) donne pour α, β, ... les valeurs

$$(4) \qquad \alpha = (aa), \qquad \beta = (ab), \qquad \gamma = (ac), \quad \ldots$$

Il en résulte pour ρ_i l'expression

$$\rho_i = (aa)a_i + (ab)b_i + (ac)c_i + \ldots$$

Calcul de ω. — Le minimum ainsi obtenu pour ω, somme des carrés des expressions ρ_i ci-dessus, se présente sous la forme d'une somme de termes qu'on peut écrire

$$\begin{aligned} &(aa)(aa)[aa] + (aa)(ab)[ab] + (aa)(ac)[ac] + \ldots \\ +&(ab)(aa)[ba] + (ab)(ab)[bb] + (ab)(ac)[bc] + \ldots \\ +&(ac)(aa)[ca] + (ac)(ab)[cb] + (ac)(ac)[cc] + \ldots \\ +&\ldots \end{aligned}$$

Or, le coefficient de (aa) dans la première ligne n'est autre, d'après (4), que

$$[aa]\alpha + [ab]\beta + [ac]\gamma + \ldots,$$

et sa valeur est 1, le coefficient de (ab) dans la deuxième ligne n'est autre que

$$[ba]\alpha + [bb]\beta + [bc]\gamma + \ldots$$

et sa valeur est zéro, les coefficients de (ac), ..., dans les lignes suivantes, sont tous nuls également en vertu du système (S).

En définitive, nous avons

$$\omega = (aa).$$

Calcul de x. — Puisque

$$\rho_i = (aa)a_i + (ab)b_i + (ac)c_i + \ldots$$

nous avons

$$(5) \qquad x = [\rho l] = (aa)[al] + (ab)[bl] + (ac)[cl] + \ldots$$

52. **Équations normales.** — On obtient, par la même méthode,

pour les inconnues y, z, ..., les valeurs

$$y = (ba)[al] + (bb)[bl] + (bc)[cl] + \ldots,$$
$$z = (ca)[al] + (cb)[bl] + (cc)[cl] + \ldots,$$

Or, dans leur ensemble, ces valeurs des inconnues, calculées d'après le principe de la moindre erreur quadratique moyenne, constituent la solution unique du système linéaire

$$(\mathrm{N}) \qquad \begin{cases} [aa]x + [ab]y + [ac]z + \ldots = [al], \\ [ba]x + [bb]y + [bc]z + \ldots = [bl], \\ [ca]x + [cb]y + [cc]z + \ldots = [cl], \\ \ldots \end{cases}$$

Et ce système s'obtient directement, à partir des équations d'erreur (E), si l'on cherche *les valeurs des inconnues qui correspondent au minimum de la somme*

$$\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2$$

des carrés des résidus.

Le principe du moindre risque d'erreur suffit, par conséquent, moyennant les hypothèses faites, *pour justifier la méthode des moindres carrés.*

Les équations (N) sont dites *équations normales* de Gauss. On doit à Gauss une démonstration très élégante du fait que, dans chacun des calculs de x, y, z, la somme des carrés ω obtenue est *effectivement minimum*.

Désignons par λ_i les valeurs des multiplicateurs ρ_i donnant à x sa valeur principale (5); nous avons, quel que soit i,

$$(6) \qquad \lambda_i = (aa)a_i + (ab)b_i + (ac)c_i + \ldots,$$

si nous désignons par ρ_i un système absolument quelconque de multiplicateurs tels que $[\varphi\rho] = x$ et vérifiant, par conséquent, les relations $[a\rho] = 1$, $[b\rho] = 0$, $[c\rho] = 0$, ..., nous avons

$$(\rho_1 - \lambda_1)a_1 + (\rho_2 - \lambda_2)a_2 + \ldots + (\rho_n - \lambda_n)a_n = 0,$$
$$(\rho_1 - \lambda_1)b_1 + (\rho_2 - \lambda_2)b_2 + \ldots + (\rho_n - \lambda_n)b_n = 0,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Ajoutons ces relations membre à membre, après multiplication par

(aa), (ab), ... ; nous obtenons, compte tenu des relations (6),

$$(\rho_1-\lambda_1)\lambda_1+(\rho_2-\lambda_2)\lambda_2+\ldots+(\rho_n-\lambda_n)\lambda_n=0,$$

résultat qu'on peut écrire

$$(7)\quad \rho_1^2+\rho_2^2+\ldots+\rho_n^2=\lambda_1^2+\lambda_2^2+\ldots+\lambda_n^2+(\rho_1-\lambda_1)^2+\ldots+(\rho_n-\lambda_n)^2$$

Ce résultat montre bien que ω *est minimum quand on a, quel que soit* i, $\rho_i=\lambda_i$. Des démonstrations entièrement semblables peuvent être données en ce qui concerne les multiplicateurs fournissant les valeurs principales des autres inconnues

53. **Poids des inconnues.** — Il résulte de ce qui précède qu'en désignant par m l'erreur quadratique moyenne de l'une quelconque des observations directes, celles-ci étant toutes d'égale précision, l'erreur quadratique moyenne sur la valeur de x deduite de la résolution des équations normales est

$$m_x = m\sqrt{[\lambda\lambda]} = m\sqrt{(\alpha\alpha)}$$

Les *poids* de plusieurs grandeurs simultanément envisagées etant inversement proportionnels aux carrés des erreurs quadratiques moyennes correspondantes, nous voyons que les quantites

$$\frac{1}{(\alpha\alpha)},\quad \frac{1}{(bb)},\quad \frac{1}{(cc)},\quad \ldots$$

sont *proportionnelles aux poids des inconnues* Chaque dénominateur est donné par la résolution, en ce qui concerne l'une des inconnues, d'un système (S) déduit du système des équations normales en remplaçant les seconds membres par zero, sauf celui de l'équation normale de même rang que l'inconnue considérée, qui est remplacé par un. Les divers systèmes (S) ainsi obtenus sont souvent appelés systèmes d'*équations aux poids*.

54 **Calcul de l'erreur à craindre.** — La formule ramenant le calcul de l'erreur quadratique moyenne relative à chaque inconnue à celui de l'erreur quadratique moyenne m, commune aux observations directes, ne résout pas le problème de l'erreur à craindre sur l'inconnue envisagée, car m n'est pas connu et doit précisément être apprécié d'après la résolution approchée des équations (1), telle que nous venons de l'envisager.

Soient ξ, η, ζ, ... les *valeurs véritables* des inconnues; en désignant par e_1, e_2, ..., e_n les *erreurs effectivement commises* dans les observations, nous avons, pour $i = 1, 2, \ldots, n$,

$$e_i = a_i\xi + b_i\eta + c_i\zeta + \ldots - l_i$$

D'autre part, si x, y, z, ... désignent les *valeurs approchées* adoptées par l'application de la méthode des moindres carrés, et ε_1, ε_2, ..., ε_n les *résidus* correspondants, nous avons

$$\varepsilon_i = a_i x + b_i y + c_i z + \ldots - l_i$$

Nous en déduisons, par soustraction,

$$(8) \qquad \varepsilon_i = e_i + a_i(x - \xi) + b_i(y - \eta) + c_i(z - \zeta) + \ldots$$

Or, si nous désignons par λ_1, λ_2, ..., λ_n les valeurs des multiplicateurs ρ_1, ρ_2, ..., ρ_n qui ont donné x, nous avons

$$[\lambda l] = x, \qquad [\lambda(e + l)] = \xi,$$

et il en résulte

$$x - \xi = -[\lambda e],$$

on a de même, en désignant par μ_1, μ_2, ..., μ_n les multiplicateurs qui ont donné y, par ν_1, ν_2, ..., ν_n ceux qui ont donné z, etc.,

$$y - \eta = -[\mu e], \qquad z - \zeta = -[\nu e],$$

Portant ces valeurs dans (8), nous obtenons

$$(9) \qquad \varepsilon_i = e_i - a_i[\lambda e] - b_i[\mu e] - c_i[\nu e] - \ldots$$

Ainsi, *les résidus sont des fonctions linéaires et homogènes des erreurs effectivement commises dans les observations.*

Formons la somme $e_1\varepsilon_1 + e_2\varepsilon_2 + \ldots + e_n\varepsilon_n$; elle s'écrit

$$(10) \qquad [e\varepsilon] = [ee] - [ae][\lambda e] - [be][\mu e] - [ce][\nu e] - \ldots;$$

or, nous avons, d'après les équations normales elles-mêmes,

$$[a\varepsilon] = 0, \qquad [b\varepsilon] = 0, \qquad [c\varepsilon] = 0, \qquad \ldots,$$

il en résulte, en formant, d'après (9), la somme $\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2$, que

$$[\varepsilon\varepsilon] = [\varepsilon e],$$

portant enfin cette valeur de $[\varepsilon e]$ dans (10), nous obtenons le résultat

important

$$[\varepsilon\varepsilon] = [ee] - [ae][\lambda e] - [be][\mu e] - [ce][\nu e] - \ldots \tag{11}$$

qui montre que *la somme des carrés des résidus est une fonction homogène du second degré des erreurs effectivement commises*.

Dans l'ignorance où nous nous trouvons des valeurs de ces erreurs, *adoptons comme valeur approchée du second membre de* (11) *sa valeur probable.*

Ce second membre se compose du terme $[ee]$ dont la valeur probable est nm^2, puis d'un certain nombre de termes contenant les carrés des erreurs, enfin d'un certain nombre de termes contenant les produits des erreurs deux à deux. Remplaçant chaque carré par sa valeur probable m^2 et chaque produit de deux erreurs par sa valeur probable zéro, nous obtenons

$$[\varepsilon\varepsilon] = nm^2 - m^2[a\lambda] - m^2[b\mu] - m^2[c\nu] - \ldots$$

c'est-à-dire, puisque $[a\lambda] = 1$, $[b\mu] = 1$, $[c\nu] = 1$, ...,

$$[\varepsilon\varepsilon] = (n - k)m^2. \tag{12}$$

Il résulte de là que m a pour valeur approchée

$$\sqrt{\frac{[\varepsilon\varepsilon]}{n-k}},$$

et que les erreurs à craindre sur x, y, z, ... (valeurs approchées des erreurs quadratiques moyennes correspondantes) sont

$$\sqrt{\frac{[\varepsilon\varepsilon](aa)}{n-k}}, \quad \sqrt{\frac{[\varepsilon\varepsilon](bb)}{n-k}}, \quad \sqrt{\frac{[\varepsilon\varepsilon](cc)}{n-k}}, \quad \ldots \tag{13}$$

On peut obtenir une expression de la somme $[\varepsilon\varepsilon]$ en fonction des coefficients des équations de condition, et, par conséquent, déterminer la somme des carrés des résidus sans devoir résoudre au préalable les équations normales en faisant intervenir ([1]) les propriétés des *formes quadratiques*.

La somme $[\varepsilon\varepsilon]$ est la valeur que prend la forme quadratique non homogène

$$F(x, y, z, \ldots) = \Sigma(a_i x + b_i y + c_i z + \ldots - l_i)^2$$

([1]) *Voir* ANDOYER, *Cours de Mécanique céleste*, t. I, Chap. IX (Gauthier-Villars).

pour les valeurs de x, y, z, . . qui annulent les dérivées F'_x, F'_y, F'_z, ... ; c'est donc, dans l'espace (x, y, z, .), le *terme constant* de l'équation de la surface du second ordre $F = 0$ par rapport à des axes parallèles aux axes initiaux et issus de son centre

Le discriminant de F restant inchangé par une translation d'axes, et les termes du premier degré disparaissant dans le cas des axes issus du centre, nous avons immédiatement

$$[zz] = \frac{D}{\Delta},$$

en désignant par Δ le déterminant des coefficients des équations normales et par D le déterminant

$$\begin{vmatrix} [aa] & [ab] & [ac] & & [al] \\ [ba] & [bb] & [bc] & & [bl] \\ & & & & . \\ [la] & [lb] & [lc] & & [ll] \end{vmatrix}$$

obtenu en bordant Δ au moyen des seconds membres des équations normales et de la somme $[ll]$

55. **Cas des observations d'inegale précision** — Supposons maintenant que les observations ayant donne les résultats L_1, L_2, . . , L_n soient de précisions inégales ; et désignons par p_1, p_2, ..., p_n les poids de ces observations, inversement proportionnels aux carrés moyens respectifs m_1, m_2, ..., m_n des erreurs commises.

Il est facile de voir que le problème se ramène à celui déjà traité relativement à des observations d'égale précision, pourvu que l'on remplace les équations de condition par celles obtenues en les multipliant respectivement par les racines carrées des poids des observations qu'elles traduisent.

Posant $a'_i = \sqrt{p_i}a_i$, $b'_i = \sqrt{p_i}b_i$, ..., nous avons $v'_i = \sqrt{p_i}v_i$, le carré moyen de l'erreur sur

$$x = \rho_1 l'_1 + \rho_2 l'_2 + \ .\ + \rho_n l'_n$$

est

$$p_1\rho_1^2 m_1^2 + p_2\rho_2^2 m_2^2 + \ .+ p_n\rho_n^2 m_n^2,$$

ou

$$m^2(\rho_1^2 + \rho_2^2 + \ldots + \rho_n^2),$$

en désignant par m^2 la valeur commune des quantités $p_i m_i^2$, c'est-à-dire le carré moyen « d'une observation de poids un ».

Les multiplicateurs ρ_i correspondant au minimum de la somme $\omega = [\rho\rho]$, tout en vérifiant les relations

$$[a'\rho] = 1, \qquad [b'\rho] = 0, \qquad [c'\rho] = 0,$$

sont des combinaisons de la forme

$$\rho_i = \alpha a'_i + \beta b'_i + \gamma c'_i + \ldots$$

les α vérifiant le système

$$(\mathrm{S}') \qquad \begin{cases} [a'a']\alpha + [a'b']\beta + [a'c']\gamma + \ldots = 1, \\ [b'a']\alpha + [b'b']\beta + [b'c']\gamma + \ldots = 0 \\ \ldots \end{cases}$$

et les valeurs principales des inconnues sont données par les équations normales

$$(\mathrm{N}') \qquad \begin{cases} [a'a']x + [a'b']y + [a'c']z + \ldots = [a'l'] \\ [b'a']x + [b'b']y + [b'c']z + \ldots = [b'l'] \\ \ldots \end{cases}$$

En remplaçant a'_i par $\sqrt{p_i}\,a_i$, b'_i par $\sqrt{p_i}\,b_i$, ..., les systèmes (S') et (N') prennent respectivement les formes

$$(\mathrm{S}') \qquad \begin{cases} [paa]\alpha + [pab]\beta + [pac]\gamma + \ldots = 1, \\ [pba]\alpha + [pbb]\beta + [pbc]\gamma + \ldots = 0, \\ \ldots \end{cases}$$

et

$$(\mathrm{N}') \qquad \begin{cases} [paa]x + [pab]y + [pac]z + \ldots = [pal], \\ [pba]x + [pbb]y + [pbc]z + \ldots = [pbl], \\ \ldots \end{cases}$$

Les équations (N') s'obtiennent en cherchant le minimum de la somme des carrés des résidus, multipliés respectivement par les poids des observations correspondantes : *c'est la forme la plus générale de la méthode des moindres carrés.*

La détermination de l'erreur à craindre se fait comme plus haut;

on a

$$m^2 = \frac{[\varepsilon\varepsilon]}{n-k}$$

et

$$m_\lambda = m\sqrt{[\lambda\lambda]}, \qquad m_\mu = m\sqrt{[\mu\mu]},$$

les quantités $\sqrt{[\lambda\lambda]}$, $\sqrt{[\mu\mu]}$, . . , étant fournies par les divers systèmes d'*équations aux poids* déduits du système (S')

CHAPITRE XI.

EXÉCUTION DES CALCULS

I. — ALGORITHME DE GAUSS POUR LA RÉSOLUTION DES ÉQUATIONS NORMALES.

56. Formation des équations normales — Supposons-nous placés dans les conditions du § 50, et examinons les opérations successives que suppose l'application de la méthode des moindres carrés telle que nous l'avons envisagée au Chapitre précédent

Nous supposons faite l'*opération préliminaire* fondamentale qui consiste à transformer les équations d'erreur initiales ayant pour inconnues X, Y, Z, ..., en les équations linéaires

$$(1)\quad \begin{cases} a_1 x + b_1 y + c_1 z + \ldots - l_1 = 0, \\ a_2 x + b_2 y + c_2 z + \ldots - l_2 = 0, \\ \ldots \\ a_n x + b_n y + c_n z + \ldots - l_n = 0, \end{cases}$$

dont les inconnues sont x, y, z, ..., les modalités de cette opération peuvent être très différentes suivant la nature de la question concrète considérée.

Nous supposerons aussi que les équations (1) ont *le même poids*, c'est-à-dire que si les observations directes effectivement faites ont des précisions différentes, les équations correspondantes ont été multipliées par les racines carrées des poids respectifs des observations, et sont devenues ainsi les équations (1).

La formation des équations normales exige alors le calcul des quantités $[aa]$, $[ab]$, ..., $[al]$, ..., qui en sont les coefficients, et auxquelles il convient d'adjoindre, si l'on veut tirer parti de l'observation finale du § 54, la quantité $[ll]$; on a ainsi $\frac{(k+1)(k+2)}{2}$ sommes à calculer, chacune étant formée de n termes.

Pour vérifier ces calculs, qui sont souvent très longs dans les applications pratiques à l'astronomie ou à la géodésie, on introduit utilement les quantités

$$s_i = a_i + b_i + c_i + \quad + l_i,$$

et l'on vérifie les $h+1$ relations telles que

$$[aa] + [ab] + \quad + [al] = [as]$$

57. **Résolution des équations normales.** — Pour la résolution des équations normales, on emploie généralement la *méthode de substitution*, et l'on utilise encore aujourd'hui la succession d'opérations suivante, avec les notations proposées par Gauss lui-même

Soient les équations normales

$$(2)\quad \begin{cases} [aa]x + [ab]y + [ac]z + \quad + [ah]w = [al], \\ [ba]x + [bb]y + [bc]z + \quad + [bh]w = [bl], \\ [ca]x + [cb]y + [cc]z + \quad + [ch]w = [cl], \\ \quad\ldots\ldots\ldots\ldots\ldots\ldots, \\ [ha]x + [hb]y + [hc]z + \quad + [hh]w = [hl] \end{cases}$$

La première équation donne, en la résolvant par rapport à la première inconnue x,

$$x = \frac{[al]}{[aa]} - \frac{[ab]}{[aa]}y - \frac{[ac]}{[aa]}z - \quad - \frac{[ah]}{[aa]}w$$

Portant cette valeur dans les autres équations, on obtient le nouveau système

$$(3)\quad \begin{cases} [bb\,1]y + [bc\,1]z + \quad + [bh\,1]w = [bl\,1], \\ [cb\,1]y + [cc\,1]z + \quad + [ch\,1]w = [cl\,1], \\ \quad\ldots\ldots\ldots\ldots\ldots\ldots, \\ [hb\,1]y + [hc\,1]z + \ldots + [hh\,1]w = [hl.1], \end{cases}$$

où, en désignant par α, β deux quelconques des lettres b, c, ..., h, l, nous posons, conformément aux notations de Gauss qui sont universellement adoptées,

$$[\alpha\beta\,1] = [\alpha\beta] - \frac{[a\alpha][a\beta]}{[aa]}.$$

La première équation (3) donne, en la résolvant par rapport à

l'inconnue correspondante y,

$$y = \frac{[bl\ 1]}{[bb\ 1]} - \frac{[bc\ 1]}{[bb\ 1]} z - \ldots - \frac{[bh\ 1]}{[bb\ 1]} w,$$

et il en résulte, en portant cette valeur dans les autres équations, le nouveau système

$$(4) \quad \begin{cases} [cc\ 2] z - \ldots - [ch\ 2] w = [cl\ 2] \\ \ldots \\ [hc\ 2] z + \ldots - [hh\ 2] w = [hl\ 2] \end{cases}$$

dans lequel nous avons posé

$$[\alpha\beta\ 2] = [\alpha\beta\ 1] - \frac{[b\alpha\ 1][b\beta\ 1]}{[bb\ 1]}.$$

Continuant ainsi de proche en proche, on parvient à l'équation unique

$$(5) \qquad [hh\ k-1] w = [hl\ k-1]$$

qui fournit la valeur de w. Remontant de proche en proche de cette valeur à celles des inconnues successivement éliminées, on achève la résolution du système.

Les calculs de formation des systèmes successifs peuvent être contrôlés, par l'emploi des sommes s, au moyen des formules

$$[bb.1] + [bc\ 1] + \ldots + [bh\ 1] + [bl\ 1] = [bs\ 1]$$
$$[cc\ 2] + \ldots + [ch\ 2] - [cl\ 2] = [cs\ 2]$$

dont l'exactitude se vérifie immédiatement de proche en proche en remontant des crochets $[\alpha\beta\ 2]$ aux crochets $[\alpha\beta.1]$ et $[\alpha\beta]$.

58. **Détermination des erreurs à craindre.** — Nous avons vu aux § 53 et 54 comment on obtient d'abord les *poids des inconnues* en résolvant les divers systèmes d'équations aux poids, déduits de celui des équations normales en remplaçant les seconds membres par 1 ou 0 suivant que l'équation a ou n'a pas le même rang que l'inconnue envisagée. Nous savons en outre que les erreurs quadratiques moyennes sont

$$m_x = m\sqrt{(aa)}, \quad \ldots, \quad m_w = m\sqrt{(hh)},$$

où

$$m = \sqrt{\frac{[\varepsilon\varepsilon]}{n-k}},$$

$[\varepsilon\varepsilon]$ désignant la *somme des carrés des résidus.*

Les divers systèmes d'équations aux poids se résolvent en suivant pas à pas les opérations déjà faites sur les équations normales. Les systèmes successifs ainsi obtenus ont, pour les inconnues, les mêmes coefficients que dans le cas des équations normales, et seuls les seconds membres diffèrent. On n'a du reste besoin, chaque fois, que d'une des inconnues : celle qui est successivement obtenue la $k^{\text{ième}}$, la $(k-1)^{\text{ième}}$, .., la première. Mais si on les calcule toutes, on a une bonne vérification des calculs, car les mineurs normalisés tels que (ab), .., qui n'interviennent pas dans le calcul des poids, sont tous obtenus deux fois.

Quant à la somme $[\varepsilon\varepsilon]$ des carrés des résidus, on peut naturellement la déduire du calcul préalable de ces résidus, lequel possède toujours son intérêt propre. Mais si l'on veut utiliser la remarque finale du § 54 et la formule

$$[\varepsilon\varepsilon] = \frac{D}{\Delta},$$

on peut obtenir une expression immédiate de cette somme.

Soient en effet D_1 et Δ_1 les déterminants analogues à D et Δ que l'on peut associer au système (3); la formation de D_1 exige le calcul de

$$[ll.1] = [ll] - \frac{[al][al]}{[aa]}.$$

On passe de D à D_1 par la suite des opérations suivantes : 1° on divise par $[aa]$ tous les éléments de la première ligne, ce qui divise D par $[aa]$; 2° on retranche des éléments de chaque ligne les quotients aussi obtenus, multipliés par $[ab]$, $[ac]$, ..., $[al]$, et cela ne change pas la valeur du déterminant $\frac{D}{[aa]}$.

Comme on passe exactement de la même manière de Δ à Δ_1, nous voyons que

$$\frac{D}{\Delta} = \frac{D_1}{\Delta_1}.$$

De la même manière, on associe au système (4) les déterminants D_2

et Δ_2 dont le rapport reste égal au rapport $\frac{D}{\Delta}$; finalement, pour le système réduit à l'équation unique (5), on a

$$D_{k-1} = \begin{vmatrix} [hh\ k-1] & [hl\ k-1] \\ [hl\ k-1] & [ll\ k-1] \end{vmatrix} \qquad \Delta_{k-1} = [hh\ k-1]$$

donc

$$\frac{D}{\Delta} = \frac{D_{k-1}}{\Delta_{k-1}} = [ll\ k-1] - \frac{[hl\ k-1][hl\ k-1]}{[hh\ k-1]},$$

ce qui revient à écrire, en prolongeant les opérations envisagées plus haut,

$$[\varepsilon\varepsilon] = [ll\ k]$$

Une étude toute pareille peut être développée en ce qui concerne le mineur normalisé (hh), inverse du poids de la dernière inconnue w. On a

$$(hh) = \frac{\delta_{(w)}}{\Delta},$$

en désignant par $\delta_{(w)}$ le coefficient de $[hh]$ dans Δ. Lorsqu'on passe du système (2) au système (3), Δ est remplacé par $\Delta_1 = \frac{\Delta}{[aa]}$, mais le coefficient de $[hh\ 1]$ dans Δ_1 est aussi déduit de $\delta_{(w)}$ en divisant par $[aa]$. De proche en proche, on arrive au résultat

$$(hh) = \frac{\delta_{k-1(w)}}{\Delta_{k-1}} = \frac{1}{\Delta_{k-1}} = \frac{1}{[hh\ k-1]},$$

de sorte que *le poids* $\frac{1}{(hh)}$ *de la dernière inconnue n'est autre que son coefficient dans la dernière équation résolvante, qui fournit la valeur de cette inconnue.*

Il en résulte pour la détermination des poids une méthode intéressante, consistant à reprendre la résolution des équations normales en plaçant au dernier rang chacune des inconnues successivement.

II — EXEMPLES NUMÉRIQUES

59 **Schéma pratique de disposition des calculs.** — Nous allons d'abord étudier la disposition des calculs, sur un exemple *purement didactique* dû à Gauss lui-même; malgré le petit nombre des équa-

tions et leur extrême simplicité, nous suivrons pas à pas le dispositif classique de la formation et de la résolution des équations.

Il s'agit du système suivant de quatre équations à trois inconnues, correspondant par hypothèse à quatre observations d'égale précision

$$
\begin{aligned}
x - y + 2z &= 3,\\
3x + 2y - 5z &= 5,\\
4x + y + 4z &= 21,\\
-x + 3y + 3z &= 14
\end{aligned}
$$

Les calculs de *formation des équations normales* sont indiqués ci-dessous, dans un tableau où les sommes s, interviennent à titre de vérification, ainsi qu'il est dit plus haut ·

NUMÉROS des équations	a	b	c	l	s
1	1	—1	2	3	5
2	3	2	—5	5	5
3	4	1	4	21	30
4	—1	3	3	14	19

NUMÉROS des équations	aa	ab	ac	al	as	bb	bc	bl	bs	cc	cl	cs	ll	ls
1	1	—1	2	3	5	1	—2	—3	—5	4	6	10	9	15
2	9	6	—15	15	15	4	—10	10	10	25	—25	—25	25	25
3	16	4	16	84	120	1	4	21	30	16	84	120	441	630
4	1	—3	—3	—14	—19	9	9	42	57	9	42	57	196	266
Sommes.	27	6	0	88	121	15	1	70	92	54	107	162	671	936

Le système des équations normales est complètement déterminé par les coefficients inscrits dans la dernière ligne du tableau ci-dessus,

avec l'écriture de l'Algèbre élémentaire, il a la forme

$$\begin{aligned} 27x + 6y \qquad &= 88 \\ 6x + 15y - \quad z &= 70 \\ y - 54z &= 107 \end{aligned}$$

Les calculs relatifs à la *résolution de ces équations*, par la formation des « équations d'élimination » successives peuvent être disposés, sans qu'il soit nécessaire d'avoir écrit les équations normales en x, y, z, et sans qu'il soit nécessaire d'écrire les équations d'élimination en y, z, puis en z, dans un tableau faisant directement suite au précédent :

	[aa]	[ab]	[ac]	[al]	[as]	[bb]	[bc]	[bl]	[bs]	[cc]	[cl]	[cs]	[ll]	[ls]
	27	6	0	88	121	15	1	70	92	54	107	162	671	936
[1]						$\frac{41}{3}$	1	$\frac{454}{9}$	$\frac{586}{9}$	54	107	162	$\frac{10373}{27}$	$\frac{14624}{27}$
[2]										$\frac{2211}{41}$	$\frac{12707}{123}$	$\frac{19340}{123}$	$\frac{24353}{123}$	$\frac{37060}{123}$
[3]													$\frac{65600}{815859}$	$\frac{65600}{815859}$

On tire des résultats de la ligne marquée [2]

$$z = \frac{[cl.2]}{[cc.2]} = \frac{12707}{6633} \text{ ou } 1,916 \text{ à un millième près}$$

on a ensuite

$$y = \frac{[bl.1]}{[bb.1]} - \frac{[bc.1]}{[bb.1]} z = \frac{2617}{737} \text{ ou } 3,551 \text{ à un millième près},$$

$$x = \frac{[al]}{[aa]} - \frac{[ab]}{[aa]} y - \frac{[ac]}{[aa]} z = \frac{49154}{19899} \text{ ou } 2,470 \text{ à un millième près}$$

Les *poids des inconnues* se déduisent, par les mêmes calculs de résolution, des divers systèmes *d'équations aux poids* formés, à partir du système des équations normales, en remplaçant respectivement les seconds membres par 1, 0, 0, par 0, 1, 0 et par 0, 0, 1. Ces calculs ne diffèrent des précédents qu'en ce qui concerne les colonnes marquées [al], [bl] et [cl]; les modifications font apparaître successi-

vement

	$[al]$	$[bl]$	$[cl]$
	1	0	0
[1]		$-\frac{6}{27}$	0
[2]			$\frac{2}{123}$

	$[al]$	$[bl]$	$[cl]$
	0	1	0
[1]		1	0
[2]			$-\frac{3}{41}$

	$[al]$	$[bl]$	$[cl]$
	0	0	1
[1]		0	1
[2]			1

On déduit du premier tableau

$$(ac) = \frac{2}{6633},$$

$$(ab) = -\frac{6}{27} \times \frac{3}{41} - \frac{3}{41} \times \frac{2}{6633} = -\frac{492}{30217},$$

et

$$(aa) = \frac{1}{27} + \frac{6}{27} \times \frac{492}{30217} = \frac{33169}{815859} \text{ ou } 0,0407 \text{ par excès}$$

— Du deuxième tableau

$$(bc) = -\frac{3}{2211},$$

$$(bb) = \frac{3}{41} + \frac{3}{2211} \times \frac{3}{41} = \frac{2214}{30217} \text{ ou } 0,0733 \text{ par excès},$$

et, à titre de vérification,

$$(ab) = -\frac{6}{27} \times \frac{2214}{30217} = -\frac{492}{30217}.$$

— Du troisième tableau

$$(cc) = \frac{41}{2211} \text{ ou } 0,0181 \text{ par excès},$$

résultat conforme à la remarque finale du paragraphe 58 d'après laquelle

$$\frac{1}{(cc)} = [cc\,2],$$

et, à titre de vérification,

$$(bc) = -\frac{3}{41} \times \frac{41}{2211} = -\frac{3}{2211},$$

$$(ac) = -\frac{6}{27} \times -\frac{3}{2211} = \frac{2}{6633}.$$

Erreurs à craindre. — La somme des carrés des résidus a pour valeur

$$[ll.3] = \frac{65600}{815859}$$

ou 0,0804 par excès. Nous avons donc ici

$$m = 0,28$$

et

$$m_x = 0,28\sqrt{(aa)} = 0,057$$

$$m_y = 0,28\sqrt{(bb)} = 0,076,$$

$$m_z = 0,28\sqrt{(cc)} = 0,038$$

Il est à peine besoin de dire que ces résultats sont donnés ici *à titre d'exemple de disposition des calculs*, et que, dans cet exemple, les observations sont *trop peu nombreuses pour qu'il soit légitime de remplacer la valeur probable d'une somme quelconque de termes par la valeur observée*.

Signalons enfin que les calculs ci-dessus ont été faits sous forme rationnelle pour mieux mettre les vérifications en évidence, mais que, dans la pratique, on les effectue le plus souvent *sous forme décimale*, avec la précision que demandent les circonstances.

60. Exemple d'application effective de la méthode. — Je dois à l'obligeance de M. Caubet, astronome-adjoint à l'Observatoire de Toulouse, la communication des calculs suivants tirés de ses travaux personnels, et relatifs à un problème, classique en Astronomie, de rectification des éléments osculateurs de l'orbite d'une petite planète.

Il s'agit de la rectification, par la méthode usuelle dite « de la variation des éléments », des éléments de l'orbite de la petite planète « 173 Ino », d'après les résultats d'observations photographiques faites aux Observatoires d'Alger et Toulouse de 1914 à 1924.

Les éléments de l'orbite étant :

l'anomalie moyenne M_0 à l'époque $t = 0$;
le moyen mouvement diurne n,
l'angle φ dont le sinus est l'excentricité de l'orbite;

l'inclinaison de l'orbite i,
la longitude du nœud ☊,
l'argument de la latitude ω;

les corrections dM_0, dn, $d\varphi$, di, d☊, $d\omega$ sont liées aux inconnues x, y, z, t, u, v par les relations

$$\begin{array}{l|c} x = dM_0, & u\cos\omega - v\sin\omega = di, \\ y = 1000\,dn, & u\sin\omega + v\cos\omega = \sin i\, d☊, \\ z = d\varphi & t + \tan\frac{i}{2}\sin i\, d☊ = d(☊ + \omega) \end{array}$$

Les formules pour la correction d'une orbite elliptique fournissent des relations linéaires entre ces inconnues et les quantités $\cos\delta.\Delta\alpha$ et $\Delta\delta$, où $\Delta\alpha$ et $\Delta\delta$ désignent les différences « observation — calcul », en ascension droite et déclinaison, constatées par rapport aux éléments initiaux à rectifier

Ces équations, linéaires en x, y, z, t, u, v, étant mises sous la forme

$$ax + by + cz + dt + eu + fv + n = 0,$$

il est fait état de 28 observations, ce qui donne un total de 56 équations d'erreur.

Les calculs de formation des équations normales, c'est-à-dire de détermination des sommes $[aa]$, $[ab]$, . . , ont été d'abord effectués pour les 28 équations d'ascension droite, ensuite pour les 28 équations de déclinaison, les équations normales définitives sont obtenues par addition des totaux partiels Les vérifications classiques

$$\begin{gathered} a + b + c + d + e + f + n = s, \\ aa + ab + ac + ad + ae + af + an = as, \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{gathered}$$

ont été constamment faites; les produits ont été calculés avec la Table de Crelle, et il n'a généralement été conservé, vu la précision poursuivie, que deux chiffres décimaux significatifs.

Les tableaux suivants contiennent les données initiales et les résultats intermédiaires complètement détaillés.

I — ASCENSION DROITE

Équations d'erreur

NUMÉROS des observations	a	b	c	d	e	f	n	s
1	1,80	−0,72	2,97	1,47	−0,17	0,09	0,7	6,14
2	1,92	−0,77	3,15	1,57	−0,20	0,09	0,7	6,46
3	1,62	−0,66	2,68	1,34	−0,23	0,04	−0,9	3,89
4	1,62	−0,66	2,68	1,34	−0,23	0,04	−0,5	4,29
5	1,58	−0,65	2,63	1,32	−0,23	0,04	−0,4	4,29
6	1,57	−0,64	2,61	1,31	−0,23	0,03	−0,3	4,55
7	1,53	−0,62	2,55	1,27	−0,22	0,03	−0,9	3,64
8	1,51	−0,61	2,53	1,26	−0,22	0,03	−0,2	4,30
9	1,50	−0,61	2,51	1,25	−0,22	0,03	−0,7	3,76
10	1,15	−0,44	2,06	1,00	−0,16	0,00	−0,2	3,41
11	0,98	0,05	0,99	1,40	0,11	0,24	−1,2	2,57
12	0,86	0,03	0,84	1,23	0,06	0,20	0,3	3,72
13	0,85	0,03	0,83	1,22	0,06	0,20	−0,1	3,09
14	0,79	0,03	0,73	1,12	0,05	0,19	−0,5	2,42
15	1,01	0,50	−1,67	1,31	−0,17	0,22	0,9	2,10
16	0,90	0,42	−1,58	1,17	−0,22	0,17	0,5	1,26
17	2,69	2,60	−1,49	1,78	0,14	0,31	2,5	8,53
18	1,27	1,87	2,55	1,45	0,07	0,03	0,2	7,44
19	1,01	1,48	1,97	1,17	0,01	0,01	0,1	5,75
20	0,91	1,73	−0,27	1,36	−0,04	0,31	0,7	4,70
21	0,79	1,49	−0,27	1,17	−0,06	0,27	1,3	4,69
22	0,77	1,46	−0,27	1,16	−0,06	0,26	2,1	5,42
23	1,51	3,55	−2,99	1,52	−0,14	0,03	−2,2	1,28
24	1,52	3,55	−3,05	1,53	−0,19	0,02	−1,2	2,18
25	1,28	2,99	−2,62	1,27	−0,18	−0,01	−1,1	1,63
26	2,12	6,09	2,93	1,60	−0,22	0,17	−2,9	9,79
27	1,21	3,51	1,93	0,98	−0,19	0,03	−2,8	4,67
28	1,30	3,75	2,02	1,04	−0,21	0,04	−1,0	6,94

Formation des équations normales

NUMÉROS des observations.	aa	ab +	ab −	ac +	ac −	ad +	ad −	ae +	ae −	af +	af −	an +	an −	as +	bb	bc +	bc −	bd +	bd −
1 ...	3,24		1,30	5,35		2,65			0,31	0,16		1,26		11,05	0,52		2,14		1,06
2 ...	3,69		1,48	6,05		3,01			0,38	0,17		1,34		12,40	0,59		2,42		1,21
3 ...	2,62		1,07	4,34		2,17			0,37	0,07			1,46	6,30	0,44		1,77		0,88
4 ...	2,62		1,07	4,34		2,17			0,37	0,07			0,81	6,95	0,44		1,77		0,88
5 ...	2,50		1,03	4,16		2,09			0,36	0,06			0,63	6,78	0,42		1,71		0,86
6 ...	2,46		1,00	4,10		2,06			0,36	0,05			0,47	6,83	0,41		1,67		0,84
7 ...	2,34		0,95	3,91		1,94			0,34	0,05			1,38	5,57	0,38		1,58		0,79
8 ...	2,28		0,92	3,82		1,90			0,33	0,05			0,30	6,49	0,37		1,54		0,77
9 ...	2,25		0,92	3,77		1,88			0,33	0,05			1,05	5,64	0,37		1,53		0,76
10 ...	1,32		0,50	2,37		1,15			0,18	0			0,23	3,92	0,19		0,91		0,44
11 ...	0,96	0,05		0,97		1,37		0,11		0,24			1,18	2,52	0,00	0,05		0,07	
12 ...	0,74	0,03		0,72		1,06		0,05		0,17		0,25		3,03	0,00	0,03		0,04	
13 ...	0,72	0,03		0,71		1,04		0,05		0,17			0,09	2,63	0,00	0,02		0,04	
14 ...	0,62	0,03		0,58		0,89		0,04		0,15			0,40	1,92	0,00	0,02		0,03	
15 ...	1,02	0,51			1,69	1,32			0,17	0,22		0,91		2,12	0,25		0,84	0,66	
16 ...	0,81	0,38			1,42	1,06			0,20	0,15		0,45		1,22	0,18		0,66	0,49	
17 ...	7,24	6,99			4,01	4,79		0,38		0,83		6,73		22,95	6,76		3,87	4,63	
18 ...	1,61	2,37		3,24		1,84		0,09		0,04		0,25		9,45	3,50	4,77		2,71	
19 ...	1,02	1,49		1,99		1,18		0,01		0,01		0,10		5,81	2,19	2,92		1,73	
20 ...	0,83	1,57			0,25	1,24			0,04	0,28		0,64		4,28	2,99		0,47	2,35	
21 ...	0,62	1,18			0,21	0,92			0,05	0,21		1,03		3,71	2,22		0,40	1,74	
22 ...	0,59	1,12			0,21	0,89			0,05	0,20		1,62		4,17	2,13		0,39	1,69	
23 ...	2,28	5,36			4,51	2,30			0,21	0,05			3,32	1,93	12,60		10,61	5,40	
24 ...	2,31	5,40			4,64	2,33			0,29	0,03			1,82	3,31	12,60		10,83	5,43	
25 ...	1,64	3,83			3,35	1,63			0,23		0,01		1,41	2,09	8,94		7,83	3,80	
26 ...	4,49	12,91		6,21		3,39			0,47	0,36			6,15	20,75	37,09	17,84		9,74	
27 ...	1,46	4,25		2,34		1,19			0,23	0,04			3,39	5,65	12,32	6,77		3,44	
28 ...	1,69	4,88		2,63		1,35			0,27	0,05			1,30	9,02	14,66	7,58		3,90	
Totaux ...	55,97	+42,14		+41 30		+50 81		— 4 81		+ 3 92		—10,81		178 49	121,96	—12 94		39 40	

Formation des équations normales (suite)

NUMÉROS des observations.	be +	be −	bf +	bf −	bn +	bn −	bs +	bs −	cc
1	0,12			0,06		0,50		4,41	8,82
2	0,15			0,07		0,54		4,97	9,92
3	0,15			0,03	0,60			2,57	7,18
4	0,15			0,03	0,33			2,83	7,18
5	0,15			0,03	0,26			2,79	6,92
6	0,15			0,02	0,19			2,78	6,81
7	0,14			0,02	0,56			2,25	6,50
8	0,13			0,02	0,12			2,62	6,40
9	0,13			0,02	0,43			2,29	6,30
10	0,07			0,00	0,09			1,50	4,24
11	0,01		0,01			0,06	0,13		0,98
12	0,00		0,01		0,01		0,11		0,71
13	0,00		0,01			0,01	0,09		0,69
14	0,00		0,01			0,02	0,07		0,53
15		0,09	0,11		0,45		1,05		2,79
16		0,09	0,07		0,21		0,57		2,50
17	0,36		0,81		6,50		22,18		2,22
18	0,13		0,06		0,37		13,91		6,50
19	0,01		0,01		0,15		8,52		3,88
20		0,07	0,54		1,21		8,13		0,07
21		0,09	0,40		1,94		6,99		0,07
22		0,09	0,38		3,07		7,91		0,07
23		0,50	0,11			7,81	4,54		8,94
24		0,67	0,07			4,26	7,73		9,30
25		0,54		0,03		3,29	4,87		6,86
26		1,34	1,04			17,66	59,62		8,58
27		0,67	0,11			9,83	16,38		3,72
28		0,79	0,15			3,75	26,03		4,08
Totaux..	− 3,09		+ 3,57		−31,24		+ 159,82		+132,76

NUMÉROS des observations.	cd +	cd −	ce +	ce −	cf +	cf −	cn +	cn −	cs +	cs −	dd	de +	de −
1	4,37			0,50	0,27		2,08		18,24		2,16		0,25
2	4,95			0,63	0,28		2,21		20,35		2,46		0,31
3	3,59			0,62	0,11			2,41	10,43		1,80		0,31
4	3,59			0,62	0,11			1,34	11,50		1,80		0,31
5	3,47			0,60	0,11			1,05	11,28		1,74		0,30
6	3,42			0,60	0,08			0,78	11,35		1,72		0,30
7	3,24			0,56	0,08			2,30	9,28		1,61		0,28
8	3,19			0,56	0,08			0,51	10,88		1,59		0,28
9	3,14			0,55	0,08			1,76	9,44		1,56		0,28
10	2,06			0,33	0,00			0,41	7,03		1,00		0,16
11	1,39		0,11		0,24			1,20	2,54		1,96	0,15	
12	1,03		0,05		0,17		0,25		2,96		1,51	0,07	
13	1,01		0,05		0,17			0,08	2,56		1,49	0,07	
14	0,82		0,05		0,14			0,37	1,77		1,28	0,06	
15		2,19	0,30			0,37		1,50		3,51	1,72		0,22
16		1,85	0,34			0,27		0,79		2,15	1,37		0,26
17		2,65		0,21		0,46		3,73		12,71	3,17	0,25	
18	3,70		0,18		0,08		0,51		18,97		2,10	0,10	
19	2,30		0,02		0,02		0,20		11,33		1,36	0,01	
20		0,37	0,01			0,08		0,19		1,27	1,85		0,05
21		0,35	0,01			0,07		0,35		1,26	1,36		0,07
22		0,31	0,02			0,07		0,37		1,47	1,35		0,07
23		4,54	0,42			0,09	6,58			3,83	2,31		0,21
24		4,67	0,58			0,06	3,66			6,65	2,34		0,29
25		3,33	0,47		0,03		2,88			4,27	1,61		0,23
26	4,69			0,64	0,50			8,50	28,68		2,56		0,35
27	1,89			0,37	0,06			5,40	9,01		0,96		0,19
28	2,10			0,42	0,08			2,02	14,02		1,08		0,22
Totaux..	+33,72		− 4,60		+ 1,22		−16,89		+ 174,50		+48,82	− 4,23	

Formation des équations normales (suite)

NUMÉROS des observations.	*df* +	*df* —	*dn* +	*dn* —	*ds* +	*ds* —	*ee*	*ef* +	*ef* —	*en* +	*en* —	*es* +	*es* —	*ff*	*fn* +	*fn* —	*fs* +	*fs* —	*nn*	*ns* +	*ns* —
1..	0,13		1,04		9,03		0,03		0,02		0,12		1,04	0,01	0,06		0,55		0,49	4,30	
2 .	0,14		1,09		10,14		0,04		0,02		0,14		1,29	0,01	0,06		0,58		0,49	4,52	
3.. ..	0,05			1,21	5,21		0,05		0,01	0,21			0,89	0,00		0,04	0,16		0,81		3,50
4 .. .	0,05			0,67	5,75		0,05		0,01	0,12			0,99	0,00		0,02	0,18		0,25		2,15
5.....	0,05			0,53	5,66		0,05		0,01	0,09			0,99	0,00		0,02	0,18		0,16		1,72
6.....	0,04			0,39	5,70		0,05		0,01	0,07			1,00	0,00		0,01	0,13		0,09		1,31
7.	0,04			1,14	4,62		0,05		0,01	0,20			0,80	0,00		0,03	0,11		0,81		3,28
8...	0,04			0,25	5,42		0,05		0,01	0,04			0,95	0,00		0,01	0,13		0,04		0,86
9	0,04			0,88	4,70		0,05		0,01	0,15			0,83	0,00		0,02	0,11		0,49		2,63
10 .	0,00			0,20	3,41		0,03		0,00	0,03			0,55	0,00		0,00	0,00		0,04		0,68
11 ..	0,34			1,68	3,60		0,01	0,03			0,13	0,28		0,06		0,29	0,62		1,44		3,08
12 .	0,25		0,37		4,33		0,00	0,01		0,02		0,21		0,04	0,06		0,70		0,09	1,06	
13 ..	0,24			0,12	3,77		0,00	0,01			0,01	0,19		0,04		0,02	0,62		0,01		0,31
14.	0,21			0,57	2,73		0,00	0,01			0,03	0,12		0,04		0,10	0,46		0,25		1,21
15.	0,29		1,18		2,75		0,03		0,04		0,15		0,36	0,05	0,20		0,46		0,81	1,89	
16	0,20		0,59		1,59		0,05		0,04		0,11		0,30	0,03	0,09		0,23		0,25	0,68	
17.. ..	0,55		4,45		15,18		0,02	0,04		0,35		1,19		0,10	0,78		2,64		6,25	21,33	
18...	0,04		0,29		10,79		0,00	0,00		0,01		0,52		0,00	0,01		0,22		0,04	1,49	
19 ..	0,01		0,12		6,73		0,00	0,00		0,00		0,06		0,00	0,00		0,06		0,01	0,58	
20	0,42		0,95		6,39		0,00		0,01		0,03		0,19	0,10	0,22		1,46		0,49	3,29	
21... ..	0,32		1,52		5,49		0,00		0,02		0,08		0,28	0,07	0,35		1,27		1,69	6,10	
22.	0,30		2,44		6,29		0,00		0,02		0,13		0,33	0,07	0,55		1,41		4,41	11,38	
23	0,05			3,34	1,95		0,02		0,00	0,31			0,18	0,00		0,07	0,04		4,84		2,82
24 .	0,03			1,84	3,34		0,04		0,00	0,23			0,41	0,00		0,02	0,04		1,44		2,62
25		0,01		1,40	2,07		0,03		0,00	0,20			0,29	0,00	0,01			0,02	1,21		1,79
26 .. .	0,27			4,64	15,66		0,05		0,04	0,64			2,1[illegible]	0,03		0,49	1,66		8,41		28,39
27 . ..	0,03			2,74	4,58		0,04		0,01	0,53			0,89	0,00		0,09	0,14		7,84		13,08
28	0,04			1,04	7,22		0,04		0,01	0,21			1,46	0,00		0,04	0,28		1,00		6,94
Totaux...	+ 4,16		— 8,60		+ 164,10		— 0,78	— 0,20		+ 2,48		—13,60		—0,65	— 1,00		—14,42		+44,15	—19,75	

II — DÉCLINAISON

Équations d'erreur

NUMÉROS des observations	*a*	*b*	*c*	*d*	*e*	*f*	*n*	*s*
1	0,35	—0,19	0,36	0,24	1,32	—0,71	—0,3	1,07
2	0,35	—0,19	0,37	0,24	1,43	—0,64	—0,1	1,46
3	0,26	—0,12	0,39	0,20	1,38	—0,24	—0,4	1,47
4	0,26	—0,12	0,39	0,20	1,38	—0,24	—0,1	1,77
5	0,26	—0,12	0,39	0,20	1,36	—0,21	0,5	2,38
6	0,25	—0,11	0,39	0,20	1,35	—0,21	0,3	2,17
7	0,25	—0,11	0,39	0,20	1,33	—0,18	0,4	2,28
8	0,25	—0,11	0,39	0,20	1,32	—0,17	—0,1	1,78
9	0,24	—0,11	0,39	0,20	1,31	—0,17	0,4	2,26
10	0,18	—0,06	0,34	0,17	1,08	0,02	—0,3	1,43
11	—0,19	—0,01	—0,21	—0,27	0,58	1,28	—0,7	0,48
12	—0,13	0,02	—0,21	—0,18	0,35	1,21	—1,0	0,06
13	—0,13	0,02	—0,21	—0,18	0,34	1,20	—0,5	0,54
14	—0,11	0,02	—0,19	—0,16	0,28	1,13	—0,5	0,47
15	—0,21	—0,08	0,28	—0,29	—0,81	1,06	0,1	0,05
16	—0,17	—0,06	0,28	—0,24	—0,96	0,75	0,5	0,10
17	0,50	0,47	—0,42	0,34	—0,72	—1,63	—1,0	—2,46
18	—0,03	—0,05	—0,06	—0,06	1,34	0,56	0,1	1,80
19	0,00	0,00	—0,02	0,01	0,97	0,74	—0,8	0,90
20	—0,20	—0,38	—0,05	—0,30	—0,18	1,34	—0,1	0,13
21	—0,15	—0,29	—0,04	—0,24	—0,27	1,17	—0,4	—0,22
22	—0,15	—0,29	—0,04	—0,23	—0,28	1,15	0,1	0,26
23	—0,12	—0,28	0,29	—0,17	—1,48	0,33	0,3	—1,13
24	—0,13	—0,30	0,32	—0,17	—1,53	0,15	0,2	—1,46
25	—0,16	—0,38	0,35	—0,17	—1,34	—0,06	—0,3	—2,06
26	0,48	1,33	0,37	0,32	1,26	—0,97	—0,9	1,89
27	0,20	0,57	0,35	0,16	1,05	—0,17	—0,6	1,56
28	0,25	0,72	0,39	0,20	1,10	—0,23	1,3	3,73

Formation des équations normales (suite).

NUMEROS des observations.	aa	ab +	ab −	ac +	ac −	ad +	ad −	ae +	ae −	af +	af −	an +	an −	as +	as −	bb +	bb −	bc +	bc −	bd +	bd −
1.	0,12		0,07	0,13		0,08		0,46			0,25		0,11	0,37			0,04		0,07		0,05
2	0,12		0,07	0,13		0,08		0,50			0,22		0,04	0,51			0,04		0,07		0,05
3.	0,07		0,03	0,10		0,05		0,36			0,06		0,10	0,38			0,01		0,05		0,02
4	0,07		0,03	0,10		0,05		0,36			0,06		0,03	0,46			0,01		0,05		0,02
5.	0,07		0,03	0,10		0,05		0,35			0,05	0,13		0,62			0,01		0,05		0,02
6	0,06		0,03	0,10		0,05		0,34			0,05	0,08		0,54			0,01		0,04		0,02
7	0,06		0,03	0,10		0,05		0,33			0,05	0,10		0,57			0,01		0,04		0,02
8.	0,06		0,03	0,10		0,05		0,33			0,04		0,03	0,45			0,01		0,04		0,02
9	0,06		0,03	0,09		0,05		0,31			0,04	0,10		0,54			0,01		0,04		0,02
10.	0,03		0,01	0,06		0,03		0,19			0,00		0,05	0,26			0,00		0,02		0,01
11	0,04		0,00	0,04		0,05			0,11		0,24	0,13			0,09		0,00		0,00		0,00
12.	0,02		0,00	0,03		0,02			0,05		0,16	0,13			0,01		0,00		0,00		0,00
13	0,02		0,00	0,03		0,02			0,04		0,16	0,07			0,07		0,00		0,00		0,00
14	0,01		0,00	0,02		0,02			0,03		0,12	0,06			0,05		0,00		0,00		0,00
15.	0,04	0,02			0,06	0,06		0,17			0,22		0,02		0,01		0,01		0,02	0,02	
16	0,03	0,01			0,05	0,04		0,16			0,13		0,09		0,02		0,00		0,02	0,01	
17	0,25	0,24			0,21	0,17			0,36		0,82		0,50		1,23		0,22		0,20	0,16	
18	0,00	0,00		0,00		0,00			0,04		0,02		0,00		0,05		0,00		0,00	0,00	
19	0,00	0,00		0,00		0,00			0,00		0,00		0,00		0,00		0,00		0,00	0,00	
20	0,04	0,08		0,01		0,06		0,04			0,27	0,02			0,03		0,14	0,02		0,11	
21	0,02	0,04		0,01		0,04		0,04			0,18	0,06		0,03			0,08	0,01		0,07	
22	0,02	0,04		0,01		0,03		0,04			0,17		0,02		0,04		0,08	0,01		0,07	
23	0,01	0,03			0,03	0,02		0,18			0,04		0,04	0,14			0,08		0,08	0,05	
24	0,02	0,04			0,04	0,02		0,20			0,02		0,03	0,19			0,09		0,10	0,05	
25	0,03	0,06			0,06	0,03		0,21		0,01		0,05		0,33			0,14		0,13	0,06	
26	0,23	0,64		0,18		0,15		0,60			0,47		0,43	0,91			1,77	0,49		0,42	
27	0,04	0,11		0,07		0,03		0,21			0,03		0,12	0,31			0,32	0,20		0,09	
28	0,06	0,18		0,10		0,05		0,28			0,06	0,33		0,93			0,15	0,28		0,14	
Totaux	1,60	− 1,13		+ 1,06		− 1,35		− 5,03		− 3,92		− 0,35		− 5,94		− 3,60		− 0,01		+ 1,00	

Formation des équations normales (suite)

NUMÉROS des observations	*be* +	*be* —	*bf* +	*bf* —	*bn* +	*bn* —	*bs* +	*bs* —	*cc*	*cd* +	*cd* —	*ce* +	*ce* —	*cf* +	*cf* —	*cn* +	*cn* —	*cs* +	*cs* —	*dd*	*de* +	*de* —
1. . . .		0,25	0,13		0,06			0,20	0,13	0,09		0,48			0,26		0,11	0,39		0,06	0,32	
2.		0,27	0,12		0,02			0,28	0,14	0,09		0,53			0,24		0,04	0,54		0,06	0,34	
3.		0,16	0,03		0,05			0,18	0,15	0,08		0,53			0,09		0,16	0,57		0,04	0,28	
4		0,16	0,03		0,01			0,21	0,15	0,08		0,53			0,09		0,04	0,69		0,04	0,28	
5. . . .		0,16	0,03			0,06		0,29	0,15	0,08		0,53			0,08	0,20		0,93		0,04	0,27	
6 . .		0,16	0,02			0,03		0,24	0,15	0,08		0,53			0,08	0,12		0,85		0,04	0,27	
7 . .		0,15	0,02			0,04		0,25	0,15	0,08		0,52			0,07	0,16		0,89		0,04	0,27	
8 . .		0,15	0,02		0,01			0,20	0,15	0,08		0,51			0,07		0,04	0,69		0,04	0,26	
9		0,14	0,02			0,04		0,25	0,15	0,08		0,51			0,07	0,16		0,88		0,04	0,26	
10.		0,06		0,00	0,02			0,09	0,12	0,06		0,37		0,01			0,10	0,49		0,03	0,18	
11.		0,01		0,01	0,01			0,00	0,04	0,06			0,12		0,27	0,15			0,10	0,07		0,16
12	0,01		0,02			0,02	0,00		0,04	0,04			0,07		0,25	0,21			0,01	0,03		0,06
13.	0,01		0,02			0,01	0,01		0,04	0,04			0,07		0,25	0,11			0,11	0,03		0,06
14 . .	0,01		0,09			0,01	0,01		0,04	0,03			0,05		0,21	0,10			0,09	0,03		0,04
15. . . .	0,06			0,08		0,01		0,00	0,08		0,08		0,23	0,30		0,03		0,01		0,08	0,23	
16	0,06			0,05		0,03		0,01	0,08		0,07		0,27	0,21		0,14		0,03		0,06	0,23	
17.		0,34		0,77		0,47		1,16	0,18		0,14	0,30		0,69		0,42		1,03		0,12		0,24
18. .		0,07		0,03		0,01		0,09	0,00	0,00			0,08		0,03		0,01		0,11	0,00		0,08
19.		0,00		0,00		0,00		0,00	0,00	0,00			0,02		0,01	0,02			0,02	0,00	0,01	
20.	0,07			0,51	0,04			0,05	0,00	0,02		0,01			0,07	0,01			0,01	0,09	0,03	
21.	0,08			0,34	0,12		0,06		0,00	0,01		0,01			0,05	0,02		0,01		0,06	0,06	
22. . .	0,08			0,33		0,03		0,08	0,00	0,01		0,01			0,05		0,00		0,01	0,05	0,06	
23.	0,41			0,09		0,08	0,32		0,08		0,05		0,43	0,10		0,09			0,33	0,03	0,25	
24 .	0,46			0,05		0,06	0,44		0,10		0,05		0,49	0,05		0,06			0,47	0,03	0,26	
25	0,51		0,02		0,11		0,78		0,12		0,06		0,47		0,02		0,11		0,72	0,03	0,23	
26. . .	1,68			1,29		1,20	2,51		0,14	0,12		0,47			0,36		0,33	0,70		0,10	0,40	
27 . . .	0,60			0,10		0,34	0,89		0,12	0,06		0,37			0,06		0,21	0,55		0,03	0,17	
28.	0,79			0,17	0,94		2,69		0,15	0,08		0,43			0,09	0,51		1,45		0,04	0,22	
Totaux .	— 2,75		— 3,32		— 1,05		+ 4,13		+2,65	+ 0,82		+ 4,34		— 1,41		+ 1,36		+ 8,72		+ 1,31	+ 4,26	

Formation des équations normales (fin).

NUMÉROS des observations	*df*		*dn*		*ds*		*ee*	*ef*		*en*		*es*		*ff*	*fn*		*fs*		*nn*	*ns*	
	+	—	+	—	+	—		+	—	+	—	+	—		+	—	+	—		+	—
1		0,17		0,07	0,26		1,74		0,94		0,40	1,41		0,50	0,21			0,76	0,09		0,32
2		0,15		0,02	0,36		2,04		0,92		0,14	2,09		0,41	0,06			0,94	0,01		0,15
3		0,05		0,08	0,29		1,90		0,33		0,55	2,03		0,06	0,10			0,35	0,16		0,59
4		0,05		0,02	0,35		1,90		0,33		0,14	2,44		0,06	0,02			0,42	0,01		0,18
5		0,04	0,10		0,48		1,85		0,29	0,68		3,24		0,04		0,11		0,50	0,25	1,19	
6		0,04	0,06		0,43		1,82		0,28	0,41		2,93		0,04		0,06		0,46	0,09	0,65	
7		0,04	0,08		0,46		1,77		0,24	0,33		3,03		0,03		0,07		0,41	0,16	0,91	
8		0,03		0,02	0,36		1,74		0,22		0,13	2,35		0,03	0,02			0,30	0,01		0,18
9		0,03	0,08		0,45		1,72		0,22	0,52		2,97		0,03		0,07		0,38	0,16	0,90	
10		0,00		0,05	0,24		1,17	0,02			0,32	1,54		0,00		0,01	0,03		0,09		0,43
11		0,34	0,19			0,13	0,34	0,74			0,41	0,28		1,64		0,90	0,61		0,49		0,34
12		0,22	0,18			0,01	0,12	0,42			0,35	0,02		1,46		1,21	0,07		1,00		0,06
13		0,22	0,09			0,10	0,12	0,41			0,17	0,18		1,44		0,60	0,65		0,25		0,27
14		0,18	0,08			0,08	0,08	0,32			0,14	0,13		1,28		0,57	0,53		0,25		0,24
15		0,31		0,03		0,01	0,66		0,86		0,08		0,04	1,12	0,11		0,05		0,01	0,01	
16		0,18		0,12		0,02	0,92		0,72		0,48		0,10	0,76	0,38		0,08		0,25	0,05	
17		0,55		0,34		0,84	0,52	1,17		0,72		1,77		2,66	1,63		4,01		1,00	2,46	
18		0,03		0,01		0,11	1,80	0,76		0,13		2,41		0,31	0,06		1,01		0,01	0,18	
19	0,01			0,01	0,01		0,94	0,72			0,78	0,87		0,55		0,59	0,67		0,64		0,72
20		0,40	0,03			0,04	0,03		0,24	0,02			0,02	1,80		0,13	0,17		0,01		0,01
21		0,28	0,10		0,05		0,07		0,32	0,11		0,06		1,37		0,47		0,26	0,16	0,09	
22		0,26		0,02		0,06	0,08		0,32		0,03		0,07	1,32	0,12		0,30		0,01	0,03	
23		0,06		0,05	0,19		2,19		0,49		0,44	1,67		0,11	0,10			0,37	0,09		0,34
24		0,03		0,03	0,25		2,34		0,23		0,31	2,23		0,02	0,03			0,22	0,04		0,29
25	0,01		0,05		0,35		1,80	0,08		0,40		2,76		0,00	0,02		0,12		0,09	0,62	
26		0,31		0,29	0,60		1,59		1,22		1,13	2,38		0,94	0,87			1,83	0,81		1,70
27		0,03		0,10	0,25		1,10		0,18		0,63	1,64		0,03	0,10			0,27	0,36		0,94
28		0,05	0,26		0,75		1,21		0,25	1,43		4,10		0,05		0,30		0,86	1,69	4,85	
Totaux	— 4,03		+ 0,04		+ 4,73		33,76	— 3,96		— 1,68		+44,30		+ 17,86	— 1,26		— 0,04		+8,19	+ 5,18	

Dans le tableau d'ensemble des calculs destinés à la résolution des équations normales, les chiffres inscrits au-dessous des *coefficients définitifs des équations normales*, sous le titre *termes soustractifs*, sont d'abord les six quantités

$$-\frac{[ab]}{[aa]}, \quad -\frac{[ac]}{[aa]}, \quad \ldots, \quad -\frac{[an]}{[aa]},$$

à utiliser à la fin des opérations pour le calcul de x, puis après la quantité $-\frac{[as]}{[aa]}$ intervenant dans les vérifications, les produits des six quantités ci-dessus par $[ab]$, termes soustractifs dans les calculs tels que

$$[bb\ 1] = [bb] - [ab]\frac{[ab]}{[aa]},$$

et ainsi de suite. La même disposition est adoptée dans toutes les rangées intitulées *termes soustractifs*.

La résolution, faite en remontant au-dessous du tableau général des coefficients numériques, conduit aux valeurs suivantes, à retenir seulement à un centième de seconde d'arc près :

$$x = 0'',14 \qquad y = 0'',36 \qquad z = 0'',21$$
$$t = -0'',42 \qquad u = -0'',02 \qquad v = -0'',01$$

Erreurs à craindre. — La somme des carrés des résidus, d'après la théorie du § 58, a pour valeur

$$[nn\ 6] = 39,15$$

et l'erreur quadratique moyenne d'une observation a pour expression

$$m = \sqrt{\frac{39,15}{55}} = 0'',84 \text{ environ.}$$

Nous ne ferons pas le calcul des poids des inconnues, en raison de sa longueur. Dans de nombreuses applications, on se contente de constater la petitesse de la quantité m calculée ci-dessus, ainsi que celle des résidus formés individuellement. Nous ne formerons pas les résidus du calcul ci-dessus, signalons seulement que 45 d'entre eux, sur 56, sont inférieurs à une seconde d'arc, c'est là une précision rarement dépassée dans les travaux astronomiques de cette nature,

aussi M Caubet a-t-il cru pouvoir s'abstenir de toute détermination d'ensemble des erreurs a craindre

Les eléments osculateurs rectifies, publiés par le *Journal des Observateurs*, et relatifs a l'equinoxe moyen de 1920, sont les suivants, pour l'époque

1924 Février 14, 0^h

$$
\begin{aligned}
M_0 &= 53^\circ 55' 54'',43, \\
n &= 781'' 2132,12 \\
\varphi &= 12^\circ 1' 3'',97, \\
\omega &= 226^\circ 7' 35'',49, \\
i &= 14^\circ 14' 47'',00, \\
\Omega &= 148^\circ 32' 52'',00
\end{aligned}
$$

CHAPITRE XII.

COMPENSATION D'OBSERVATIONS CONDITIONNELLES

61. **Calcul des corrections.** — Reprenons le problème examiné aux § 20 et 46, et admettons, comme dans toute cette troisième Partie, que les observations directes ne comportent que des erreurs accidentelles très petites.

Désignons par $e_1, e_2, \ldots, e_n$ les erreurs effectivement commises $U_1 - L_1, U_2 - L_2, \ldots, U_n - L_n$. Posons

$$(1)\qquad \left\{\begin{array}{l} g_1(L_1, L_2, \ldots, L_n) = w_1, \\ g_2(L_1, L_2, \ldots, L_n) = w_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ g_p(L_1, L_2, \ldots, L_n) = w_p \end{array}\right.$$

et

$$\begin{array}{llll} \dfrac{\partial g_1}{\partial L_1} = A_1, & \dfrac{\partial g_1}{\partial L_2} = A_2, & \ldots, & \dfrac{\partial g_1}{\partial L_n} = A_n, \\ \dfrac{\partial g_2}{\partial L_1} = B_1, & \dfrac{\partial g_2}{\partial L_2} = B_2, & \ldots, & \dfrac{\partial g_2}{\partial L_n} = B_n, \\ \ldots, & \ldots, & \ldots, & \ldots, \\ \dfrac{\partial g_p}{\partial L_1} = H_1, & \dfrac{\partial g_p}{\partial L_2} = H_2, & \ldots, & \dfrac{\partial g_p}{\partial L_n} = H_n \end{array}$$

écrivons, en les limitant aux termes du premier degré en $e_1, e_2, \ldots, e_n$, les développements des équations de condition

$$(2)\qquad \left\{\begin{array}{l} A_1 e_1 + A_2 e_2 + \ldots + A_n e_n + w_1 = 0, \\ B_1 e_1 + B_2 e_2 + \ldots + B_n e_n + w_2 = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ H_1 e_1 + H_2 e_2 + \ldots + H_n e_n + w_p = 0 \end{array}\right.$$

Proposons-nous de déterminer les corrections $\lambda_1, \lambda_2, \ldots, \lambda_n$ par la double condition que, d'abord, les *valeurs compensées* $u_1 = L_1 + \lambda_1$, $u_2 = L_2 + \lambda_2, \ldots, u_n = L_n + \lambda_n$ vérifient les équations de condi-

tion, ce qui revient à écrire

$$(3)\quad \begin{cases} A_1\lambda_1 + A_2\lambda_2 + \dots + A_n\lambda_n + w_1 = 0, \\ B_1\lambda_1 + B_2\lambda_2 + \dots + B_n\lambda_n + w_2 = 0, \\ \dots\dots\dots\dots\dots\dots, \\ H_1\lambda_1 + H_2\lambda_2 + \dots + H_n\lambda_n + w_p = 0, \end{cases}$$

ensuite que *la valeur probable de chacun des carrés des erreurs commises en adoptant les valeurs compensées soit minimum* (*principe du moindre risque d'erreur*).

Ces erreurs sont $U_1 - u_1 = e_1 - \lambda_1$, $U_2 - u_2 = e_2 - \lambda_2$, ..., $U_n - u_n = e_n - \lambda_n$.

Les corrections, très petites, sont des fonctions des quantités w_1, w_2, ..., w_p; réduisant ces fonctions, qui doivent évidemment être nulles si tous les w sont nuls, à leur développement limité à sa partie linéaire, nous écrivons

$$(4)\quad \begin{cases} \lambda_1 = \alpha_1 w_1 + \beta_1 w_2 + \dots + \eta_1 w_p, \\ \lambda_2 = \alpha_2 w_1 + \beta_2 w_2 + \dots + \eta_2 w_p, \\ \dots\dots\dots\dots\dots\dots, \\ \lambda_n = \alpha_n w_1 + \beta_n w_2 + \dots + \eta_n w_p. \end{cases}$$

Nous devons donc assurer le minimum de la valeur probable de chacune des quantités

$$\varepsilon_i^2 = (e_i - \alpha_i w_1 - \beta_i w_2 - \dots - \eta_i w_p)^2,$$

dont la première s'écrit, moyennant les formules (2),

$$(5)\quad \varepsilon_1^2 = [(1 + A_1\alpha_1 + B_1\beta_1 + \dots + H_1\eta_1)e_1 + (A_2\alpha_1 + B_2\beta_1 + \dots + H_2\eta_1)e_2 + \dots]^2.$$

Supposons d'abord les observations d'égale précision, et soit m l'erreur quadratique moyenne relative à l'une quelconque de ces observations. L'expression (5) est une fonction homogène et du second degré des erreurs effectivement commises, et nous aurons sa valeur probable, d'après les hypothèses faites, en remplaçant tous les carrés e_i^2 par m^2, et tous les produits $e_i e_j$ par zéro.

La valeur probable est donc $m^2 P_1$, avec

$$P_1 = (1 + A_1\alpha_1 + B_1\beta_1 + \dots + H_1\eta_1)^2 + (A_2\alpha_1 + B_2\beta_1 + \dots + H_2\eta_1)^2 + \dots

Egalant à zero les dérivées $\frac{\partial P_1}{\partial \alpha_1}, \frac{\partial P_1}{\partial \beta_1}, \ldots, \frac{\partial P_1}{\partial \eta_1}$, et utilisant les notations

$$[AA] = A_1^2 + A_2^2 + \ldots + A_n^2,$$
$$[AB] = A_1 B_1 + A_2 B_2 + \ldots + A_n B_n,$$
$$\ldots\ldots\ldots,$$

nous obtenons le système linéaire

$$\begin{aligned}
&[AA]\alpha_1 + [AB]\beta_1 + \ldots + [AH]\eta_1 + A_1 = 0,\\
&[BA]\alpha_1 + [BB]\beta_1 + \ldots + [BH]\eta_1 + B_1 = 0,\\
&\ldots\ldots\ldots,\\
&[HA]\alpha_1 + [HB]\beta_1 + \ldots + [HH]\eta_1 + H_1 = 0,
\end{aligned}$$

qui, adjoint à la première relation (4), donne λ_1 par la relation

$$(6)\qquad \begin{vmatrix} [AA] & [AB] & \ldots & [AH] & A_1 \\ [BA] & [BB] & \ldots & [BH] & B_1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ [HA] & [HB] & \ldots & [HH] & H_1 \\ \omega_1 & \omega_2 & \ldots & \omega_p & -\lambda_1 \end{vmatrix} = 0$$

Des équations toutes semblables donnent $\lambda_2, \lambda_3, \ldots, \lambda_n$. Or, il est facile de voir que *les valeurs ainsi obtenues pour les corrections sont conformes à la méthode des moindres carrés*.

En effet, si l'on cherche à réaliser le minimum de la somme des carrés des corrections, on est conduit à trouver les valeurs des λ_i qui vérifient les équations (3) et qui, de plus, sont telles que la relation

$$\lambda_1 d\lambda_1 + \lambda_2 d\lambda_2 + \ldots + \lambda_n d\lambda_n = 0$$

soit la conséquence des relations

$$\begin{aligned}
&A_1 d\lambda_1 + A_2 d\lambda_2 + \ldots + A_n d\lambda_n = 0,\\
&B_1 d\lambda_1 + B_2 d\lambda_2 + \ldots + B_n d\lambda_n = 0,\\
&\ldots\ldots\ldots,\\
&H_1 d\lambda_1 + H_2 d\lambda_2 + \ldots + H_n d\lambda_n = 0.
\end{aligned}$$

Ces valeurs doivent d'abord être des combinaisons linéaires de la forme

$$(7)\qquad \left\{\begin{aligned}
\lambda_1 &= k_1 A_1 + k_2 B_1 + \ldots + k_p H_1,\\
\lambda_2 &= k_1 A_2 + k_2 B_2 + \ldots + k_p H_2,\\
&\ldots\ldots\ldots,\\
\lambda_n &= k_1 A_n + k_2 B_n + \ldots + k_p H_n,
\end{aligned}\right.$$

ensuite vérifier le système (3). Donc les coefficients $k_1, k_2, \ldots, k_p$, que Gauss appelle les *corrélatifs*, doivent vérifier le système

$$(8)\quad \begin{cases} [AA]k_1 + [AB]k_2 + \ldots + [AH]k_p + w_1 = 0, \\ [BA]k_1 + [BB]k_2 + \ldots + [BH]k_p + w_2 = 0, \\ \ldots \\ [HA]k_1 + [HB]k_2 + \ldots + [HH]k_p + w_p = 0. \end{cases}$$

On peut ainsi déterminer λ_1 en éliminant les corrélatifs entre les équations (8) et la première équation (7). On obtient ainsi l'équation

$$\begin{vmatrix} [AA] & [AB] & \ldots & [AH] & w_1 \\ [BA] & [BB] & \ldots & [BH] & w_2 \\ \ldots & & & & \\ [HA] & [HB] & \ldots & [HH] & w_p \\ A_1 & B_1 & \ldots & H_1 & -\lambda_1 \end{vmatrix} = 0$$

qui est identique à l'équation (6).

Les équations (8) sont appelées *équations normales* ou *équations aux corrélatifs*. On peut leur appliquer la méthode de résolution étudiée au Chapitre XI. Leur nombre est égal à *celui des équations de condition*; dans la méthode des observations indirectes, le nombre des équations normales à résoudre est égal à celui des inconnues X, Y, Z, ..., c'est-à-dire, somme toute, à *la différence entre le nombre des observations faites et le nombre des équations de condition*. Donc, au point de vue du nombre d'équations normales à résoudre, la méthode de compensation est la plus simple si le nombre des équations de condition est inférieur à la moitié du nombre des observations.

62. Détermination de l'erreur à craindre. — Proposons-nous de déterminer la valeur du carré moyen de l'erreur $\varepsilon_1 = U_1 - u_1$, c'est le minimum du produit $m^2 P_1$ envisagé au paragraphe précédent.

Or, soit ζ_1 une variable d'homogénéité, par l'emploi de laquelle P_1 s'écrit

$$(\zeta_1 + A_1\alpha_1 + B_1\beta_1 + \ldots + H_1\eta_1)^2 + (A_2\alpha_1 + B_2\beta_1 + \ldots + H_2\eta_1)^2 + \ldots$$

nous avons, par la théorie des fonctions homogènes,

$$P_1 = \frac{1}{2}\left[\alpha_1 \frac{\partial P_1}{\partial \alpha_1} + \beta_1 \frac{\partial P_1}{\partial \beta_1} + \ldots + \eta_1 \frac{\partial P_1}{\partial \eta_1} + \zeta_1 \frac{\partial P_1}{\partial \zeta_1}\right],$$

donc, si les p premières dérivées sont nulles, le minimum de P_1 s'écrit

$$(9) \qquad \frac{1}{2}\frac{\partial P_1}{\partial \zeta_1} = 1 + A_1\alpha_1 + B_1\beta_1 + \dots + H_1\eta_1,$$

et l'élimination de $\alpha_1, \beta_1, \dots, \eta_1$ entre les équations qui déterminent ces coefficients et la relation (9) donne, en désignant par μ_1^2 le carré moyen de l'erreur $U_1 - u_1$

$$\begin{vmatrix} [AA] & [AB] & \cdots & [AH] & A_1 \\ [BA] & [BB] & \cdots & [BH] & B_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ [HA] & [HB] & \cdots & [HH] & H_1 \\ A_1 & B_1 & \cdots & H_1 & 1 - \frac{\mu_1^2}{m^2} \end{vmatrix} = 0$$

Si donc nous désignons par Δ le déterminant du système des équations normales, nous avons

$$\mu_1^2 = m^2 \frac{D_1}{\Delta},$$

où

$$D_1 = \begin{vmatrix} & & & & A_1 \\ & & & & B_1 \\ & & \Delta & & \cdots \\ & & & & H_1 \\ A_1 & B_1 & \cdots & H_1 & 1 \end{vmatrix}$$

Ainsi se trouvent déterminées les valeurs quadratiques moyennes des erreurs sur les valeurs compensées, *en supposant connu le carré moyen de l'erreur relative aux observations directes*.

Le problème consiste maintenant, puisque, dans la réalité, m^2 n'est pas connu, à en déterminer une évaluation fondée sur la connaissance des quantités $w_1, w_2, \dots, w_n$.

Nous allons, pour cela, chercher la *valeur probable de la somme des carrés des corrections* $\lambda_1, \lambda_2, \dots, \lambda_n$.

Cette somme, formée des termes

$$(\alpha_1 w_1 + \beta_1 w_2 + \dots + \eta_1 w_p)^2 + \dots$$

est, d'après les relations (2), une forme quadratique par rapport aux e_i. Nous aurons sa valeur moyenne en multipliant par m^2 la

somme des coefficients des termes carrés, cette somme n'est pas autre chose que la somme des racines de l'équation en S relative à cette forme, et cette remarque est applicable à la recherche de la valeur moyenne d'une forme quadratique quelconque par rapport aux e_i.

Remarquons que nous avons la relation

$$(10) \qquad \lambda_1^2+\lambda_2^2+\dots+\lambda_n^2 = e_1\lambda_1+e_2\lambda_2+\dots+e_n\lambda_n;$$

en effet, d'après (7), nous écrivons le second membre

$$e_1\lambda_1+e_2\lambda_2+\dots+e_n\lambda_n = k_1\Sigma A_i e_i+k_2\Sigma B_i e_i+\dots+k_p\Sigma H_i e_i,$$

c'est-à-dire, d'après (2),

$$= k_1 w_1 + k_2 w_2 + \dots + k_p w_p;$$

d'autre part, le premier membre s'écrit

$$\lambda_1^2+\lambda_2^2+\dots+\lambda_n^2 = k_1\Sigma A_i\lambda_i+k_2\Sigma B_i\lambda_i+\dots+k_p\Sigma H_i\lambda_i,$$

c'est-à-dire, d'après (3),

$$= k_1 w_1 + k_2 w_2 + \dots + k_p w_p.$$

Il résulte de (10) que

$$(11) \qquad \Sigma e_i^2 = \Sigma\lambda_i^2+\Sigma(e_i-\lambda_i)^2,$$

puisque cette relation revient à

$$2\Sigma\lambda_i^2 = 2\Sigma e_i\lambda_i.$$

L'équation en S relative à la forme $\Sigma\lambda_i^2$ envisagée ci-dessus s'obtient en annulant le discriminant de la forme

$$F = \Sigma\lambda_i^2 - S\Sigma e_i^2,$$

qu'on écrit encore

$$F = (1-S)\Sigma\lambda_i^2 - S\Sigma(e_i-\lambda_i)^2.$$

Or, les λ_i peuvent, d'après (7), s'exprimer linéairement en fonction homogène de p d'entre eux, et la somme $\Sigma\lambda_i^2$ est une somme de p carrés de combinaisons linéaires, homogènes, indépendantes, des e_i

$$\Sigma\lambda_i^2 = \xi_1^2+\xi_2^2+\dots+\xi_p^2.$$

D'autre part, les différences telles que $e_i-\lambda_i$ vérifient les relations obtenues en retranchant membre à membre les équations de même

rang des systèmes (2) et (3), et par conséquent s'expriment linéairement en fonction de $n-p$ d'entre elles, la somme de leurs carrés peut s'écrire

$$\Sigma(e_i - r_i)^2 = \omega_1^2 + \omega_2^2 + \dots + \omega_{n-p}^2,$$

les ω étant des combinaisons linéaires, homogènes, indépendantes des e_i. L'ensemble des ξ et des ω forme un système de combinaisons indépendantes des e_i, car la somme de leurs carrés est, d'après (11), la forme Σe_i^2 dont le discriminant est 1. Donc nous pouvons envisager le changement de variables passant de $e_1, e_2, \dots, e_n$ aux quantités $\xi_1, \xi_2, \dots, \xi_p, \omega_1, \omega_2, \dots, \omega_{n-p}$, le discriminant de F par rapport aux e_i et celui par rapport aux nouvelles variables sont proportionnels, donc s'annulent pour les mêmes valeurs de S.

L'équation en S s'écrit immédiatement, avec

$$F = (1-S)(\xi_1^2 + \xi_2^2 + \dots + \xi_p^2) - S(\omega_1^2 + \omega_2^2 + \dots + \omega_{n-p}^2),$$

$$(1-S)^p(-S)^{n-p} = 0,$$

elle admet p racines égales à 1, $n-p$ égales à zéro, donc la somme des racines est p, et *la valeur probable de la somme des carrés des corrections est pm^2*. Il en résulte que la valeur probable de la somme des carrés des erreurs commises sur les valeurs compensées est $(n-p)m^2$.

Nous pouvons obtenir une valeur approchée de m^2 en égalant la valeur probable de la somme des carrés des corrections à sa valeur effective, nous obtenons ainsi

$$m^2 = \frac{1}{p}[rr]$$

et nous avons par conséquent

$$\mu_i^2 = \frac{D_i}{\Delta}\frac{[rr]}{p}.$$

63. **Exemple**. — Comme nous l'avons indiqué plus haut, les problèmes de la *géodésie* constituent les exemples les plus importants parmi les problèmes de compensation d'observations conditionnelles.

Considérons par exemple, l'opération de mesure des angles d'un triangle sphérique, la somme de ces angles doit être égale à 180 degrés plus « l'excès sphérique » du triangle, qui est le quotient de son aire,

connue avec précision par ailleurs, par le carré du rayon de la sphère, et que l'on réduit pratiquement en secondes, soit ε cet excès.

Imaginons que l'on ait trouvé, pour les angles A, B, C d'un tel triangle, les valeurs α, β, γ, et soit h la différence

$$180^0 + \varepsilon - (\alpha + \beta + \gamma)$$

entre la valeur réelle de la somme des angles et celle de la somme de leurs mesures. Si l'on désigne par $\alpha + x$, $\beta + y$, $\gamma + z$ les valeurs compensées, on a l'unique équation de condition

$$x + y + z = h$$

les corrélatifs se réduisent à un, les trois corrections x, y, z sont égales, et leur valeur commune est $\frac{h}{3}$. Les valeurs compensées sont donc $\alpha + \frac{h}{3}$, $\beta + \frac{h}{3}$, $\gamma + \frac{h}{3}$.

La somme des carrés des corrections est $\frac{h^2}{3}$, donc le carré de l'erreur commise dans la mesure de chaque angle a pour valeur moyenne approchée $\frac{h^2}{3}$, puisque ici $p = 1$, la somme des carrés des erreurs à craindre sur les valeurs compensées est $\frac{2}{3} h^2$, donc chacune a pour valeur $\frac{\sqrt{2}}{3} h$.

64. **Cas des observations d'inégale précision.** — Supposons que les observations ayant donné les résultats L_1, L_2, ..., L_n soient d'inégale précision, désignons par g_1, g_2, ..., g_n leurs poids, liés aux erreurs quadratiques moyennes m_1, m_2, ..., m_n par les relations

$$(12) \qquad g_1 m_1^2 = g_2 m_2^2 = \ldots = g_n m_n^2$$

Avec les notations des § 61 et 62, la valeur probable du carré de l'erreur $\varepsilon_1 = e_1 - \lambda_1$ a pour expression

$$\begin{aligned} & m_1^2(1 + A_1\alpha_1 + B_1\beta_1 + \ldots + H_1\eta_1)^2 \\ + & m_2^2(\quad A_2\alpha_1 + B_2\beta_1 + \ldots + H_2\eta_1)^2 \\ + & \ldots \ldots \ldots \end{aligned}$$

ou, en désignant par μ^2 la valeur commune des produits (12),

$$\mu^2\left[\begin{array}{l}\frac{1}{g_1}(1+A_1\alpha_1+B_1\beta_1+\dots+H_1\eta_1)^2\\ +\frac{1}{g_2}(\quad A_2\alpha_1+B_2\beta_1+\dots+H_2\eta_1)^2\\ +\dots\end{array}\right]$$

En égalant à zéro les dérivées du polynome entre crochets par rapport à $\alpha_1, \beta_1, \dots, \eta_1$, nous obtenons le système linéaire

$$\begin{aligned}
&\left[\frac{AA}{g}\right]\alpha_1+\left[\frac{AB}{g}\right]\beta_1+\dots+\left[\frac{AH}{g}\right]\eta_1+\frac{A_1}{g_1}=0,\\
&\left[\frac{BA}{g}\right]\alpha_1+\left[\frac{BB}{g}\right]\beta_1+\dots+\left[\frac{BH}{g}\right]\eta_1+\frac{B_1}{g_1}=0,\\
&\dots\\
&\left[\frac{HA}{g}\right]\alpha_1+\left[\frac{HB}{g}\right]\beta_1+\dots+\left[\frac{HH}{g}\right]\eta_1+\frac{H_1}{g_1}=0,
\end{aligned}$$

qui, adjoint à la première relation (4), donne λ_1 par la relation

$$(13)\qquad \left|\begin{array}{cccc|c} & & & & \frac{A_1}{g_1}\\ & & \Delta & & \frac{B_1}{g_1}\\ & & & & \vdots\\ & & & & \frac{H_1}{g_1}\\ \hline \omega_1 & \omega_2 & \dots & \omega_p & -\lambda_1\end{array}\right| = 0$$

où Δ désigne le déterminant des coefficients de $\alpha_1, \beta_1, \dots, \eta_1$.

Or, une étude toute semblable à celle du § 61, sur le minimum de la somme $\Sigma g_i\lambda_i^2$ des carrés des corrections multipliés par les poids des observations correspondantes, conduit à remplacer le système (7) par le système

$$(14)\qquad \left\{\begin{aligned}
\lambda_1 &= k_1\frac{A_1}{g_1}+k_2\frac{B_1}{g_1}+\dots+k_p\frac{H_1}{g_1},\\
\lambda_2 &= k_1\frac{A_2}{g_2}+k_2\frac{B_2}{g_2}+\dots+k_p\frac{H_2}{g_2},\\
&\dots\dots\dots,\\
\lambda_n &= k_1\frac{A_n}{g_n}+k_2\frac{B_n}{g_n}+\dots+k_p\frac{H_n}{g_n},
\end{aligned}\right.$$

de sorte que les corrélatifs $k_1, k_2, \dots, k_n$ vérifient les équations normales

$$(15)\quad \left\{\begin{array}{l}
\left[\dfrac{AA}{g}\right]k_1+\left[\dfrac{AB}{g}\right]k_2+\cdots+\left[\dfrac{AH}{g}\right]k_p+w_1=0,\\
\left[\dfrac{BA}{g}\right]k_1+\left[\dfrac{BB}{g}\right]k_2+\cdots+\left[\dfrac{BH}{g}\right]k_p+w_2=0,\\
\qquad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\
\left[\dfrac{HA}{g}\right]k_1+\left[\dfrac{HB}{g}\right]k_2+\cdots+\left[\dfrac{HH}{g}\right]k_p+w_p=0
\end{array}\right.$$

L'élimination des corrélatifs entre les relations (15) et la première relation (14) donne immédiatement λ_1 par une relation identique à la relation (13). Ainsi la méthode des moindres carrés se trouve étendue au cas des observations conditionnelles d'inégale précision. Restent à déterminer les erreurs à craindre.

Le carré moyen μ_1^2 de l'erreur ε_1 est le minimum du produit $\mu^2 P_1$, avec

$$\begin{aligned}
P_1 = {} & \frac{1}{g_1}(1+A_1\alpha_1+B_1\beta_1+\cdots+H_1\eta_1)^2\\
& +\frac{1}{g_2}(\quad A_2\alpha_1+B_2\beta_1+\cdots+H_2\eta_1)^2\\
& +\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,
\end{aligned}$$

une étude toute semblable à celle du § 56 montre que

$$\frac{\mu_1^2}{\mu^2}=\frac{1}{g_1}(1+A_1\alpha_1+B_1\beta_1+\cdots+H_1\eta_1),$$

de sorte que l'élimination de $\alpha_1, \beta_1, \dots, \eta_1$ entre cette relation et les équations qui déterminent ces coefficients donne immédiatement

$$\mu_1^2=\mu^2\frac{D_1}{\Delta},$$

avec

$$D_1=\left|\begin{array}{cccc|c}
 & & & & \dfrac{A_1}{g_1}\\
 & & \Delta & & \dfrac{B_1}{g_1}\\
 & & & & \vdots\\
 & & & & \dfrac{H_1}{g_1}\\
\hline
\dfrac{A_1}{g_1} & \dfrac{B_1}{g_1} & \cdots & \dfrac{H_1}{g_1} & \dfrac{1}{g_1}
\end{array}\right|$$

Les erreurs quadratiques moyennes relatives aux erreurs compensées se trouvent donc exprimées en fonction de μ^2, nous allons déterminer une valeur approchée de μ^2 en adoptant comme valeur approchée de la valeur probable de la somme

$$\Omega = g_1\lambda_1^2 + g_2\lambda_2^2 + \dots + g_n\lambda_n^2$$

la valeur effectivement obtenue par l'emploi de la méthode des moindres carrés

Cette somme, qui s'écrit

$$\begin{aligned} & g_1(\alpha_1 w_1 + \beta_1 w_2 + \dots + \eta_1 w_p)^2 \\ + & g_2(\alpha_2 w_1 + \beta_2 w_2 + \dots + \eta_2 w_p)^2 \\ + & \dots \end{aligned}$$

est, d'après les relations (2), une forme quadratique par rapport aux e_i. Nous obtiendrons sa valeur moyenne en remplaçant partout les carrés e_i^2 par les valeurs moyennes $m_i^2 = \frac{\mu^2}{g_i}$ correspondantes, et les produits $e_i e_j$ par zéro. Le résultat obtenu n'est pas autre chose que le produit de μ^2 par la somme des racines du discriminant de la forme

$$F = \Omega - \lambda[g_1 e_1^2 + g_2 e_2^2 + \dots + g_n e_n^2]$$

Les relations (10) et (11) du § 61 prennent, dans le cas actuel, la forme suivante : on a d'abord

$$(16)\qquad g_1\lambda_1^2 + g_2\lambda_2^2 + \dots + g_n\lambda_n^2 = g_1\lambda_1 e_1 + g_2\lambda_2 e_2 + \dots + g_n\lambda_n e_n,$$

en effet, d'après (14), le second membre s'écrit

$$g_1\lambda_1 e_1 + \dots + g_n\lambda_n e_n = k_1\Sigma A_i e_i + k_2\Sigma B_i e_i + \dots + k_p\Sigma H_i e_i,$$

c'est-à-dire, d'après (2),

$$-k_1 w_1 - k_2 w_2 - \dots - k_p w_p,$$

d'autre part, toujours d'après (14), on écrit le premier membre

$$g_1\lambda_1^2 + \dots + g_n\lambda_n^2 = k_1\Sigma A_i\lambda_i + k_2\Sigma B_i\lambda_i + \dots + k_p\Sigma H_i\lambda_i,$$

c'est-à-dire, d'après (3),

$$-k_1 w_1 - k_2 w_2 - \dots - k_p w_p.$$

Il résulte immédiatement de (16) que

$$\Sigma g_i e_i^2 = \Sigma g_i \lambda_i^2 + \Sigma g_i (e_i - \lambda_i)^2 \tag{17}$$

Revenons, dans ces conditions, à la forme quadratique F, on peut l'écrire

$$F = (1-\lambda)\Sigma g_i \lambda_i^2 - \lambda \Sigma g_i (e_i - \lambda_i)^2$$

Or, les λ_i peuvent, d'après (14), s'exprimer linéairement en fonction homogène de p d'entre eux, et la somme $\Sigma g_i \lambda_i^2$ est une somme de p carrés de combinaisons linéaires, homogènes, indépendantes, des e_i

$$\Sigma g_i \lambda_i^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_p^2$$

De même, on écrit comme au § 62

$$\Sigma g_i (e_i - \lambda_i)^2 = \omega_1^2 + \omega_2^2 + \dots + \omega_{n-p}^2$$

Les variables $\xi_1, \xi_2, \dots, \xi_p, \omega_1, \omega_2, \dots, \omega_{n-p}$ forment un système de formes linéaires indépendantes, car le discriminant de la forme constituée par la somme de leurs carrés est égal, d'après (17), au produit $g_1 g_2 \dots g_n$, donc est différent de zéro.

Le discriminant de la forme F peut donc se former à partir de son expression

$$(1-\lambda)(\xi_1^2 + \xi_2^2 + \dots + \xi_p^2) - \lambda(\omega_1^2 + \dots + \omega_{n-p}^2),$$

en fonction des nouvelles variables, il a pour racines p fois $\lambda = 1$, n fois $\lambda = 0$. Donc la somme des racines est p, et *la valeur probable de la somme $\Sigma g_i \lambda_i^2$ est $p\mu^2$*.

Nous pouvons donc écrire pour μ^2 (carré moyen de l'erreur d'une observation de poids un) la valeur approchée

$$\mu^2 = \frac{1}{p}[g\lambda\lambda],$$

et nous avons par conséquent

$$\mu_i^2 = \frac{D_i}{\Delta} \frac{[g\lambda\lambda]}{p}.$$

Signalons, pour terminer cette étude, la possibilité d'un contrôle simple du calcul de la somme $[g\lambda\lambda]$. Nous avons, moyennant les relations (14),

$$\begin{aligned}[g\lambda\lambda] = \; & \lambda_1[k_1 A_1 + k_2 B_1 + \dots + k_p H_1] \\ & + \lambda_2[k_1 A_2 + k_2 B_2 + \dots + k_p H_2] \\ & + \dots\end{aligned}$$

ou encore

$$\begin{aligned}[g\lambda r] = {} & k_1[\lambda_1 A_1 + \lambda_2 A_2 + \dots + \lambda_n A_n] \\ & + k_2[\lambda_1 B_1 + \lambda_2 B_2 + \dots + \lambda_n B_n] \\ & + \dots\end{aligned}$$

ou enfin, moyennant les relations (3),

$$[grr] = -[kw]$$

Reprenons, pour indiquer un exemple simple relatif aux observations d'inégale précision, le cas du *triangle* envisagé au § 63, et supposons que les valeurs α, β, γ sont respectivement les moyennes de n_1, n_2 et n_3 observations

Nous cherchons les valeurs des corrections x, y, z telles que l'on ait

$$x + y + z = h,$$

et que la somme

$$n_1 x^2 + n_2 y^2 + n_3 z^2$$

soit minimum

Posons

$$x = \frac{k}{n_1}, \qquad y = \frac{k}{n_2}, \qquad z = \frac{k}{n_3},$$

nous avons

$$k = \frac{h}{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}},$$

d'où les corrections

$$x = h\frac{\frac{1}{n_1}}{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}},$$

$$y = h\frac{\frac{1}{n_2}}{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}},$$

$$z = h\frac{\frac{1}{n_3}}{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}}$$

La somme

$$n_1 x^2 + n_2 y^2 + n_3 z^2$$

a pour valeur $\frac{h^2}{\frac{n_1}{1} + \frac{n_2}{1} + \frac{n_3}{1}}$, nous avons par ailleurs, dans le cas

actuel, $A_1 = A_2 = A_3 = 1$,

$$\Delta = \left[\frac{AA}{g}\right] = \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}$$

et

$$D_1 = \begin{vmatrix} \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} & \frac{1}{n_1} \\ \frac{1}{n_1} & \frac{1}{n_1} \end{vmatrix} = \frac{1}{n_1}\left[\frac{1}{n_2} + \frac{1}{n_3}\right],$$

il en résulte pour le carré moyen μ_1^2 de l'erreur sur la valeur $\alpha + x$ l'expression

$$\mu_1^2 = h^2 \frac{\frac{1}{n_1}\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}{\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}\right)^2}$$

65 **Exemple d'application effective de la méthode** — Nous allons étudier, pour terminer, la compensation effective d'un quadrilatère géodésique, sur un exemple tiré de la Méridienne de France (Cours de Géodésie du Service Géographique de l'Armée)

Dans la pratique des opérations géodésiques, on mesure tous les angles possibles, donc, dans le cas du quadrilatère ABCD (*fig* 4),

Fig 4

les huit angles indiqués, or, supposant donnés les sommets A et B, le quadrilatère peut être déterminé par les coordonnées des sommets C et D soit quatre conditions Entre les angles α_1, α_2, β_1, β_2, γ_1, γ_2, δ_1, δ_2, on doit donc avoir *quatre équations de condition.*

Désignant par ε_1, ε_2, ε_3, ε_4 les excès sphériques respectifs des triangles BCD, ACD, ABD, ABC, nous avons les relations suivantes, où les valeurs α_1, α_2, désignent les *résultats de mesure*, et où les

corrections sont x_1, x_2, y_1, y_2, z_1, z_2, u_1, u_2.

Triangle	ABD	$\alpha_1 + x_1 + \alpha_2 + x_2 + \beta_1 + y_1 + \delta_2 + u_2 = 200^{\mathrm{G}} + \varepsilon_3$,
»	BCD	$\beta_2 + y_2 + \gamma_1 + z_1 + \gamma_2 + z_2 + \delta_1 + u_1 = 200^{\mathrm{G}} + \varepsilon_1$,
»	ACD	$\alpha_1 + x_1 + \gamma_2 + z_2 + \delta_1 + u_1 + \delta_2 + u_2 = 200^{\mathrm{G}} + \varepsilon_2$,
»	ABC	$\alpha_2 + x_2 + \beta_1 + y_1 + \beta_2 + y_2 + \gamma_1 + z_1 = 200^{\mathrm{G}} + \varepsilon_4$.

Mais la quatrième de ces relations est une conséquence des trois autres, car on a $\varepsilon_1 + \varepsilon_3 = \varepsilon_2 + \varepsilon_4$; il faut donc trouver une quatrième équation de condition, à substituer par exemple à la dernière des équations ci-dessus.

Or, dans le triangle (plan ou sphérique) BCD, l'application du théorème de Jean de Ceva généralisé donne la relation suivante, écrite sans convention de signe, et qui exprime la concourance des directions AC, AB, AD :

$$\frac{\sin \widehat{ABC}}{\sin \widehat{ABD}} \, \frac{\sin \widehat{ACD}}{\sin \widehat{ACB}} \, \frac{\sin \widehat{ADB}}{\sin \widehat{ADC}} = 1.$$

Nous écrirons donc la quatrième équation de condition sous la forme

$$\begin{aligned} &\log \sin(\beta_1 + y_1 + \beta_2 + y_2) + \log \sin(\gamma_2 + z_2) + \log \sin(\delta_2 + u_2) \\ &\quad - \log \sin(\beta_1 + y_1) - \log \sin(\gamma_1 + z_1) - \log \sin(\delta_1 + u_1 + \delta_2 + u_2) = 0. \end{aligned}$$

Dans la pratique, on écrit chaque fois

$$\log \sin(\beta_1 + \beta_2 + y_1 + y_2) = \log \sin(\beta_1 + \beta_2) + \Delta(y_1 + y_2),$$

où Δ est la différence tabulaire pour une seconde, en supposant les corrections exprimées en secondes.

Abordons maintenant l'exemple numérique suivant.

Résultats de mesure des angles.

$\alpha_1 = 55^{\mathrm{G}}\ 1590'',608$

$\alpha_2 = 48\ \ 2303,321$

$\beta_1 = 34\ \ 3979,389$

$\beta_2 = 34\ \ 0829,366$

$\gamma_1 = 83\ \ 2890,870$

$\gamma_2 = 47\ \ 6399,207$

$\delta_1 = 34\ \ 9881,735$

$\delta_2 = 62\ \ 2135,173$

Excès sphériques

$$\varepsilon_1 = 3''938$$
$$\varepsilon_2 = 3''880$$
$$\varepsilon_3 = 5,580$$
$$\varepsilon_4 = 5,638$$

Formation des équations de condition — Nous avons successivement

$$\alpha_1 + \alpha_2 + \beta_1 + \delta_2 = 200'\,0008''\,491,$$
$$\beta_2 + \gamma_1 + \gamma_2 + \delta_1 = 200\;0001\;178,$$
$$\alpha_1 + \gamma_2 + \gamma_1 + \delta_2 = 200\;0006\;723,$$

il en résulte les « équations aux angles »

$$x_1 + x_2 + y_1 + u_2 + 2''911 = 0$$
$$y_2 + z_1 + z_2 + u_1 - 2\;760 = 0,$$
$$x_1 + z_2 + u_1 + u_2 + 2\;843 = 0$$

D'autre part, les logarithmes intervenant dans la quatrième équation de condition étant pris à neuf décimales, conformement à la précision habituelle des calculs géodésiques, cette équation s'écrit

$$\begin{aligned} 0,7689\,y_1 - 0,3684\,y_2 + 0,1833\,z_1 - 0,7348\,z_2 & \\ + 0,0300\,u_1 - 0,4303\,u_2 + 3,721 & = 0, \end{aligned}$$

toutes réductions effectuées

La méthode consistant à déterminer les valeurs des inconnues x_1, x_2, y_1, y_2, z_1, z_2, u_1, u_2 qui rendent minimum la somme

$$x_1^2 + x_2^2 + y_1^2 + y_2^2 + z_1^2 + z_2^2 + u_1^2 + u_2^2,$$

tout en vérifiant les quatre équations de condition, nous posons

$$\begin{aligned} x_1 &= k_1 + k_3, \\ x_2 &= k_1, \\ y_1 &= k_1 + 0,7689\,k_4, \\ y_2 &= k_2 - 0,3684\,k_4, \\ z_1 &= k_2 + 0,1833\,k_4, \\ z_2 &= k_2 + k_3 - 0,7348\,k_4, \\ u_1 &= k_2 + k_3 + 0,0300\,k_4, \\ u_2 &= k_1 + k_3 - 0,4303\,k_4, \end{aligned}$$

k_1, k_2, k_3, k_4 désignant les *corrélatifs*

Portant ces valeurs dans les equations de condition, nous obtenons les *équations normales*

$$4\,k_1 \qquad\qquad + 2\,k_3 + 0,3386\,k_4 - 2,911 = 0,$$
$$4\,k_2 + 2\,k_3 - 0,8899\,k_4 - 2,760 = 0$$
$$2\,k_1 + 2\,k_2 + 4\,k_3 - 1,1351\,k_4 - 2,843 = 0$$
$$0,3386\,k_1 - 0,8899\,k_2 - 1,1351\,k_3 + 1,48652\,k_4 - 3,721 = 0$$

La resolution par la methode classique donne lieu au tableau de calculs suivant

		TERMES soustractifs	[1]	TERMES soustractifs	[2]	TERMES soustractifs	[3]
aa	4,0000						
ab							
ac	2,0000	−0,50000					
ad	0,3386	−0,08465					
al	2,911	−0,72775					
bb	4,0000		4,0000				
bc	2,0000		2,0000	−0,50000			
bd	−0,8899		−0,8899	0,22248			
bl	−2,760		−2,7600	0,69000			
cc	4,0000	−1,0000	3,0000	−1,0000	2,0000		
cd	−1,1351	−0,16930	−1,3044	0,44495	−0,85945	0,42972	
cl	2,843	−1,45550	1,38750	1,38000	2,76750	−1,38375	
dd	1,48652	−0,02866	1,45786	−0,19798	1,25988	−0,36932	0,89056
dl	3,721	−0,24642	3,47458	−0,61403	2,86055	1,18926	4,04981

La derniere equation d'elimination

$$0,89056\,k_4 - 4,04981 = 0$$

donne

$$k_4 = -4,54749,$$

il en résulte de proche en proche

$$k_3 = 0,42972\,k_4 - 1,38375 = -3,33790,$$
$$k_2 = -0,5\,k_3 + 0,22248\,k_4 + 0,6900 = 1,34722,$$

et

$$k_1 = -0,5\,k_3 - 0,08465\,k_4 - 0,72775 = 1,32615$$

Les corrections ont par conséquent les valeurs

$$x_1 = -2'',01175,$$
$$x_2 = 1,32615,$$
$$y_1 = -2,17042,$$
$$y_2 = 3,02252,$$
$$z_1 = 0,51367,$$
$$z_2 = 1,35082,$$
$$u_1 = -2,12710,$$
$$u_2 = -0,05497,$$

et les angles compensés sont

$$\alpha_2 + x_2 = 55^{g}\ 1588'',596,$$
$$\alpha_2 + x_2 = 48\ 2304,647,$$
$$\beta_1 + y_1 = 34\ 3977,219,$$
$$\beta_2 + y_2 = 34\ 0832,389,$$
$$\gamma_1 + z_1 = 83\ 2891,384,$$
$$\gamma_2 + z_2 = 47\ 6400,558,$$
$$\delta_1 + u_1 = 34\ 9879,608,$$
$$\delta_2 + u_2 = 62\ 2135,118$$

Valeurs de la fonction $\Theta(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}\,dx$

	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7
00	00000	11246	22270	32863	42839	52050	60386	67780
01	00113	11358	22379	32966	42935	52138	60464	67849
02	00226	11470	22487	33069	43031	52226	60543	67918
03	00338	11581	22595	33172	43127	52313	60621	67987
04	00451	11693	22703	33275	43223	52401	60700	68056
05	00564	11805	22812	33377	43319	52488	60778	68124
06	00677	11916	22920	33480	43415	52576	60856	68193
07	00790	12028	23028	33583	43510	52663	60934	68262
08	00903	12139	23136	33686	43606	52750	61012	68330
09	01016	12251	23244	33788	43701	52837	61090	68398
10	01128	12362	23352	33891	43797	52924	61168	68466
11	01241	12474	23460	33993	43892	53011	61246	68535
12	01354	12585	23568	34096	43987	53098	61323	68603
13	01467	12697	23676	34198	44083	53185	61401	68671
14	01580	12808	23784	34300	44178	53272	61478	68738
15	01692	12920	23891	34403	44273	53358	61556	68806
16	01805	13031	23999	34505	44368	53445	61633	68874
17	01918	13142	24107	34607	44463	53531	61710	68941
18	02031	13253	24214	34709	44557	53617	61787	69009
19	02144	13365	24322	34811	44652	53704	61864	69076
20	02256	13476	24430	34913	44747	53790	61941	69143
21	02369	13587	24537	35014	44841	53876	62018	69210
22	02482	13698	24644	35116	44936	53962	62095	69277
23	02595	13809	24752	35218	45030	54048	62171	69344
24	02707	13921	24860	35319	45124	54134	62248	69411
25	02820	14032	24966	35421	45219	54219	62324	69478
26	02933	14143	25074	35523	45313	54305	62400	69545
27	03046	14254	25181	35624	45407	54390	62477	69611
28	03159	14365	25288	35725	45501	54476	62552	69678
29	03271	14476	25395	35827	45595	54561	62629	69744
30	03384	14587	25502	35928	45689	54646	62705	69810
31	03497	14698	25609	36029	45782	54732	62780	69877
32	03610	14808	25716	36130	45876	54817	62856	69943
33	03722	14919	25823	36231	45970	54902	62932	70009
34	03835	15030	25930	36332	46063	54986	63007	70074
35	03948	15141	26037	36433	46157	55071	63083	70140
36	04060	15252	26144	36534	46250	55156	63158	70206
37	04173	15363	26250	36635	46343	55241	63233	70272
38	04286	15473	26357	36735	46436	55325	63308	70337
39	04398	15584	26463	36836	46529	55410	63384	70402
40	04511	15695	26570	36936	46622	55494	63459	70468
41	04624	15805	26676	37037	46715	55578	63533	70533
42	04736	15916	26783	37137	46808	55662	63608	70598
43	04849	16026	26889	37238	46901	55746	63683	70663
44	04962	16137	26996	37338	46994	55830	63757	70728
45	05074	16247	27102	37438	47086	55914	63832	70793
46	05187	16358	27208	37538	47179	55998	63906	70858
47	05299	16468	27314	37638	47271	56082	63981	70922
48	05412	16579	27420	37738	47364	56165	64055	70987
49	05525	16689	27527	37838	47456	56249	64129	71051

Valeurs de la fonction $\Theta(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7
50	05637	16800	27633	37938	47548	56332	64203	71116
51	05750	16910	27739	38038	47640	56416	64277	71180
52	05862	17020	27844	38138	47732	56499	64351	71244
53	05975	17130	27950	38237	47824	56582	64424	71308
54	06087	17240	28056	38337	47916	56665	64498	71372
55	06200	17351	28162	38436	48008	56748	64571	71436
56	06312	17461	28268	38536	48100	56831	64644	71500
57	06424	17571	28373	38635	48191	56914	64718	71563
58	06537	17681	28479	38735	48283	56996	64791	71627
59	06650	17791	28584	38834	48374	57079	64863	71690
60	06762	17901	28690	38933	48466	57162	64938	71753
61	06874	18011	28795	39032	48557	57244	65011	71817
62	06988	18121	28901	39131	48648	57326	65083	71880
63	07100	18231	29006	39230	48739	57408	65156	71942
64	07212	18341	29111	39329	48830	57491	65229	72006
65	07324	18451	29217	39428	48921	57572	65301	72069
66	07436	18560	29322	39526	49012	57655	65374	72132
67	07549	18670	29427	39625	49103	57736	65446	72194
68	07661	18780	29532	39724	49193	57818	65518	72257
69	07773	18890	29637	39822	49284	57900	65590	72320
70	07886	19000	29742	39921	49374	57982	65662	72382
71	07998	19109	29847	40019	49465	58063	65733	72444
72	08110	19218	29951	40117	49555	58144	65807	72507
73	08223	19328	30056	40215	49645	58226	65878	72569
74	08335	19437	30161	40314	49735	58307	65950	72631
75	08447	19547	30266	40412	49826	58388	66022	72693
76	08559	19656	30370	40510	49916	58469	66093	72755
77	08671	19766	30475	40608	50006	58550	66165	72816
78	08783	19875	30579	40705	50096	58631	66236	72878
79	08896	19984	30684	40803	50185	58712	66307	72940
80	09008	20093	30788	40901	50275	58792	66378	73001
81	09120	20203	30892	40998	50364	58873	66449	73062
82	09232	20312	30996	41096	50454	58953	66520	73124
83	09344	20421	31101	41194	50543	59034	66591	73185
84	09456	20530	31205	41291	50633	59114	66662	73246
85	09568	20639	31309	41388	50722	59194	66732	73307
86	09680	20748	31413	41486	50811	59274	66803	73368
87	09792	20857	31517	41583	50900	59354	66873	73429
88	09904	20966	31621	41680	50989	59434	66944	73489
89	10016	21075	31725	41777	51078	59514	67014	73550
90	10128	21184	31828	41874	51167	59594	67084	73610
91	10240	21293	31932	41971	51256	59673	67154	73671
92	10352	21402	32036	42067	51344	59753	67224	73731
93	10464	21510	32139	42164	51433	59832	67294	73791
94	10576	21619	32243	42261	51521	59912	67364	73851
95	10687	21728	32346	42357	51609	59991	67433	73911
96	10799	21836	32450	42444	51698	60070	67503	73971
97	10911	21945	32553	42540	51786	60149	67572	74031
98	11023	22053	32656	42637	51874	60228	67642	74091
99	11135	22162	32759	42733	51962	60307	67711	74151

Valeurs de la fonction $\Theta(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2}\, dx$

	0,8	0,9	1,0	1,1	1,2	1,3	1,4	1,5
00	74210	79691	84270	88020	91031	93401	95228	96610
01	74269	79741	84311	88054	91058	93422	95244	96622
02	74329	79791	84353	88088	91085	93442	95260	96634
03	74388	79841	84394	88121	91111	93463	95276	96646
04	74447	79891	84435	88154	91138	93484	95292	96658
05	74506	79941	84477	88188	91164	93504	95307	96669
06	74565	79990	84518	88221	91191	93525	95323	96681
07	74624	80040	84559	88254	91217	93545	95339	96693
08	74683	80089	84599	88287	91243	93566	95354	96704
09	74742	80139	84640	88320	91269	93586	95370	96716
10	74800	80188	84681	88353	91295	93606	95385	96728
11	74859	80237	84722	88386	91322	93627	95401	96739
12	74917	80287	84762	88419	91348	93647	95416	96751
13	74975	80336	84803	88452	91374	93667	95431	96762
14	75034	80385	84843	88484	91399	93687	95447	96773
15	75092	80434	84883	88517	91425	93707	95462	96785
16	75150	80482	84924	88549	91451	93727	95477	96796
17	75208	80531	84964	88582	91477	93747	95492	96808
18	75266	80580	85004	88614	91502	93767	95507	96819
19	75323	80628	85044	88647	91528	93787	95523	96830
20	75381	80677	85084	88679	91553	93806	95538	96841
21	75439	80725	85124	88711	91579	93826	95553	96852
22	75496	80773	85163	88743	91604	93846	95568	96864
23	75553	80822	85203	88775	91629	93866	95582	96875
24	75611	80870	85243	88807	91655	93885	95597	96886
25	75668	80918	85282	88839	91680	93905	95612	96897
26	75725	80966	85321	88871	91705	93924	95627	96908
27	75782	81013	85361	88902	91730	93943	95641	96919
28	75839	81061	85400	88934	91755	93963	95656	96930
29	75896	81109	85440	88966	91780	93982	95671	96941
30	75952	81156	85478	88997	91805	94001	95686	96952
31	76009	81204	85517	89028	91830	94021	95700	96962
32	76065	81251	85556	89060	91855	94040	95715	96973
33	76122	81298	85595	89091	91879	94059	95729	96484
34	76178	81345	85634	89122	91904	94078	95744	96995
35	76234	81393	85673	89153	91928	94097	95758	97005
36	76291	81440	85711	89184	91953	94116	95773	97016
37	76347	81487	85750	89216	91977	94135	95787	97027
38	76403	81534	85788	89247	92001	94154	95801	97037
39	76458	81580	85827	89277	92026	94173	95815	97048
40	76514	81627	85865	89308	92050	94191	95830	97059
41	76570	81674	85903	89339	92075	94210	95844	97069
42	76625	81720	85941	89370	92099	94229	95858	97080
43	76681	81767	85979	89400	92123	94247	95872	97090
44	76736	81813	86017	89431	92147	94266	95886	97100
45	76792	81859	86055	89461	92171	94284	95900	97111
46	76847	81905	86093	89492	92195	94303	95914	97121
47	76902	81951	86131	89522	92219	94321	95928	97131
48	76957	81997	86168	89552	92243	94340	95942	97142
49	77012	82043	86206	89582	92266	94358	95956	97152

Valeurs de la fonction $\Theta(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

	0,8	0,9	1,0	1,1	1,2	1,3	1,4	1,5
50	77067	82089	86244	89612	92290	94376	95969	97162
51	77121	82135	86281	89642	92314	94394	95983	97172
52	77176	82180	86318	89672	92337	94413	95997	97183
53	77231	82226	86356	89702	92361	94431	96011	97193
54	77285	82271	86393	89732	92384	94449	96024	97203
55	77340	82317	86430	89762	92407	94467	96038	97213
56	77394	82362	86467	89791	92431	94485	96051	97223
57	77448	82407	86504	89821	92454	94503	96065	97233
58	77502	82452	86541	89851	92477	94520	96078	97243
59	77556	82497	86578	89880	92500	94538	96092	97253
60	77610	82542	86614	89910	92524	94556	96105	97263
61	77664	82587	86651	89939	92547	94574	96119	97273
62	77717	82632	86688	89968	92570	94592	96132	97282
63	77771	82677	86724	89997	92592	94609	96145	97292
64	77825	82721	86760	90027	92615	94627	96159	97302
65	77878	82766	86797	90056	92638	94644	96172	97312
66	77931	82810	86833	90085	92661	94662	96185	97322
67	77985	82854	86869	90114	92684	94679	96198	97331
68	78038	82899	86905	90142	92706	94697	96211	97341
69	78091	82943	86941	90171	92729	94714	96224	97351
70	78144	82987	86977	90200	92751	94731	96237	97360
71	78197	83031	87013	90229	92774	94748	96250	97370
72	78250	83075	87049	90257	92796	94766	96263	97380
73	78302	83119	87085	90286	92819	94783	96276	97390
74	78355	83162	87120	90314	92841	94800	96289	97399
75	78407	83206	87156	90343	92863	94817	96302	97408
76	78460	83250	87191	90371	92885	94834	96315	97417
77	78512	83293	87227	90399	92907	94851	96327	97427
78	78564	83337	87262	90427	92929	94868	96340	97436
79	78617	83380	87297	90456	92951	94885	96353	97445
80	78669	83423	87333	90484	92973	94902	96365	97455
81	78721	83466	87368	90512	92995	94918	96378	97464
82	78773	83509	87403	90540	93017	94935	96391	97473
83	78824	83552	87438	90568	93039	94952	96403	97482
84	78876	83595	87473	90595	93061	94968	96416	97492
85	78928	83638	87507	90623	93082	94985	96428	97501
86	78979	83681	87542	90651	93104	95002	96440	97510
87	79031	83723	87577	90678	93125	95018	96453	97519
88	79082	83766	87611	90706	93147	95034	96465	97528
89	79133	83808	87646	90733	93168	95051	96477	97537
90	79184	83851	87680	90761	93190	95067	96490	97546
91	79235	83893	87715	90788	93211	95084	96502	97555
92	79286	83935	87750	90815	93232	95100	96514	97564
93	79337	83977	87783	90843	93254	95116	96526	97573
94	79388	84019	87817	90870	93275	95132	96538	97582
95	79439	84061	87851	90897	93296	95148	96551	97591
96	79489	84103	87885	90924	93317	95165	96563	97600
97	79540	84145	87919	90951	93338	95181	96575	97608
98	79590	84187	87953	90978	93359	95197	96587	97617
99.	79640	84228	87987	91005	93380	95213	96590	97626

Valeurs de la fonction $\Theta(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx$

x	$\Theta(x)$	x	$\Theta(x)$	x	$\Theta(x)$	x	$\Theta(x)$
1,60	97635	2,00	99532	2,40	99931	2,80	999925
1,61	97721	2,01	99552	2,41	99933	2,81	999929
1,62	97804	2,02	99572	2,42	99938	2,82	999933
1,63	97884	2,03	99591	2,43	99941	2,83	999937
1,64	97962	2,04	99609	2,44	99944	2,84	999941
1,65	98038	2,05	99626	2,45	99947	2,85	999944
1,66	98110	2,06	99642	2,46	99950	2,86	999948
1,67	98181	2,07	99658	2,47	99952	2,87	999951
1,68	98249	2,08	99673	2,48	99955	2,88	999954
1,69	98315	2,09	99688	2,49	99957	2,89	999956
1,70	98379	2,10	99702	2,50	99959	2,90	999959
1,71	98441	2,11	99715	2,51	99961	2,91	999961
1,72	98500	2,12	99728	2,52	99963	2,92	999964
1,73	98558	2,13	99741	2,53	99965	2,93	999966
1,74	98608	2,14	99752	2,54	99967	2,94	999968
1,75	98667	2,15	99764	2,55	99969	2,95	999970
1,76	98719	2,16	99773	2,56	99970	2,96	999972
1,77	98769	2,17	99783	2,57	99972	2,97	999973
1,78	98817	2,18	99793	2,58	99974	2,98	999975
1,79	98864	2,19	99805	2,59	99975	2,99	999976
1,80	98909	2,20	99814	2,60	99976	3,00	999978
1,81	98952	2,21	99822	2,61	99978	3,01	999979
1,82	98994	2,22	99831	2,62	99979	3,02	999980
1,83	99035	2,23	99839	2,63	99980	3,03	999982
1,84	99074	2,24	99846	2,64	99981	3,04	999983
1,85	99111	2,25	99854	2,65	99982	3,05	999984
1,86	99147	2,26	99861	2,66	99983	3,06	999985
1,87	99182	2,27	99867	2,67	99984	3,07	999986
1,88	99216	2,28	99874	2,68	99985	3,08	999987
1,89	99248	2,29	99880	2,69	99986	3,09	999988
1,90	99279	2,30	99886	2,70	999866	3,10	999988
1,91	99309	2,31	99891	2,71	999873	3,11	999989
1,92	99338	2,32	99896	2,72	999880	3,12	999990
1,93	99366	2,33	99902	2,73	999887	3,13	999990
1,94	99392	2,34	99906	2,74	999893	3,14	999991
1,95	99418	2,35	99911	2,75	999899	3,15	9999916
1,96	99443	2,36	99915	2,76	999905	3,16	9999921
1,97	99466	2,37	99920	2,77	999910	3,17	9999926
1,98	99489	2,38	99924	2,78	999916	3,18	9999931
1,99	99511	2,39	99927	2,79	999920	3,19	9999936

FIN DU FASCULE II DU TOME I.

TABLE DES MATIÈRES.

Pages
VII

PREMIÈRE PARTIE.

Principes généraux.

CHAPITRE I. — *Probabilité des causes* ... 1

1. Formule de Bayes. — 2. Étude d'un exemple. — 3. Importance des probabilités *a priori*. — 4. Cas des probabilités continues. — 5. Étude d'un problème plus général.

CHAPITRE II. — *Lois de probabilité à une variable* ... 10

I. *Généralités* ... 10

6. Détermination d'une loi de probabilité. — 7. Valeurs moyennes et moments. — 8. L'écart quadratique moyen, son importance.

II. *Le problème des moments* ... 15

9. Le problème algébrique d'ordre n. — 10. Propriétés des polynomes $Q_n(x)$. — 11. Problème général des moments.

III. *Fonction caractéristique d'une loi de probabilité* ... 22

12. Définition et exemples. — 13. La fonction caractéristique détermine la loi de probabilité. — 14. Exemples. — 15. Fonction caractéristique d'une somme de plusieurs variables éventuelles.

CHAPITRE III. — *Le problème général des erreurs d'observation* ... 33

16. Erreurs accidentelles et erreurs systématiques. — 17. La loi de Gauss. — 18. Combinaison d'observations directes. — 19. Combinaison d'observations indirectes. — 20. Compensation d'observations conditionnelles.

DEUXIÈME PARTIE.

La loi de Gauss.

CHAPITRE IV. — *Propriétés générales de la loi de Gauss* ... 41

21. La fonction Θ, la courbe en cloche. — 22. Moments et fonction carac-

Pages.

téristique de la loi de Gauss. — 23. Erreur absolue moyenne ; erreur médiane, erreur quadratique moyenne — 24 Reduction d'une loi de probabilité. — 25. La loi de Gauss et la théorie des épreuves répétées — 26 Combinaison linéaire d'erreurs vérifiant la loi de Gauss. — 27. Moyenne arithmétique des mesures, poids des observations

CHAPITRE V — *Le principe de la moyenne arithmétique* 55

28 Remarques générales sur la justification de la loi de Gauss — 29. La démonstration de Gauss. — 30 Objections à la démonstration précédente — 31 Discussion de Poincaré

CHAPITRE VI — *Justification de la loi de Gauss par la méthode des moments.* 64

32 La théorie des erreurs partielles — 33 Indication fournie par la théorie des épreuves répétées. — 34 Le problème des moments pour une loi de probabilité variable — 35. Le problème général des moments pour la fonction de Gauss — 36 Théorème limite fondamental

CHAPITRE VII — *Justification de la loi de Gauss par la méthode des fonctions caractéristiques* 75

37 Fonction caractéristique d'une loi de probabilite variable 38 Loi résultante réduite — 39 Limite de la loi résultante réduite. — 40 Discussion des hypothèses faites

CHAPITRE VIII — *Combinaison des observations vérifiant la loi de Gauss* . 83

I *Combinaison d'observations directes* 83

41 Observations directes d'égale précision — 42 Observations directes d'inegale précision. — 43 Examen du même problème au point de vue de la statistique mathématique

II *Combinaison des observations indirectes ou conditionnelles* 91

44. Observations indirectes. — 45. Interpolation parabolique. — 46. Observations conditionnelles

TROISIÈME PARTIE.

Méthode des moindres carrés.

CHAPITRE IX — *Combinaison d'observations directes*. 99

47 Principe du moindre risque d'erreur. — 48. Observations directes d'inégale précision — 49. Calcul de l'erreur à craindre.

CHAPITRE X — *Combinaison d'observations indirectes* 104

50. Équations d'erreur linéaires. — 51. Détermination de x par le minimum de l'erreur quadratique moyenne. — 52. Equations normales. — 53. Poids

Pages

des inconnues. — 54. Calcul de l'erreur à craindre — 55. Cas des observations d'inégale précision

CHAPITRE XI. — *Exécution des calculs* 115

I. *Algorithme de Gauss pour la résolution des équations normales...* 115

56. Formation des équations normales. — 57. Résolution des équations normales. — 58. Détermination des erreurs à craindre.

II. *Exemples numériques.* 119

59. Schéma pratique de disposition des calculs. — 60. Exemple d'application effective de la méthode

CHAPITRE XII. — *Compensation d'observations conditionnelles* 135

61. Calcul des corrections. — 62. Détermination de l'erreur à craindre. — 63. Exemple. — 64. Cas des observations d'inégale précision. — 65. Exemple d'application effective de la méthode

Table des valeurs de la fonction $\Theta(x)$ 153

FIN DE LA TABLE DES MATIÈRES.

Collection de Monographies

SUR LA

Théorie des Fonctions

PUBLIÉE SOUS LA DIRECTION

de M. Émile BOREL

Membre de l'Institut

VOLUMES IN-8 (25×16) SE VENDANT SÉPARÉMENT

Leçons sur la théorie des fonctions (Éléments et principes de la théorie des ensembles; applications à la théorie des fonctions), par ÉMILE BOREL. 3ᵉ édition .... 25 fr

Leçons sur les fonctions entières, par ÉMILE BOREL. 2ᵉ édition.. 25 fr

Leçons sur les séries divergentes, par ÉMILE BOREL. 2ᵉ édition, entièrement revue, avec le concours de GEORGES BOULIGAND, professeur à la Faculté des Sciences de Poitiers. ... . 40 fr

Leçons sur les séries à termes positifs, professées au Collège de France par ÉMILE BOREL, recueillies et rédigées par *Robert d'Adhémar* .... . 25 fr

Leçons sur les fonctions méromorphes, professées au Collège de France par ÉMILE BOREL, recueillies et rédigées par *Ludovic Zoretti*......... 25 fr

Leçons sur l'intégration et la recherche des fonctions primitives, professées au Collège de France, par HENRI LEBESGUE. 2ᵉ édition. ..... 60 fr

Leçons sur les fonctions de variables réelles et les développements en séries de polynomes, professées à l'École Normale supérieure par ÉMILE BOREL et rédigées par *Maurice Fréchet*, avec des Notes par PAUL PAINLEVÉ et HENRI LEBESGUE. 3ᵉ édition .... 25 fr

Leçons sur les fonctions discontinues, professées au Collège de France par RENÉ BAIRE, rédigées par *A. Denjoy* .... 25 fr

Le calcul des résidus et ses applications à la théorie des fonctions, par ERNEST LINDELÖF.. 25 fr

Leçons sur les séries trigonométriques, professées au Collège de France par HENRI LEBESGUE.......... .. .. (*Sous presse*.)

Leçons sur les fonctions définies par les équations différentielles du premier ordre, professées au Collège de France par PIERRE BOUTROUX; avec une *Note* de PAUL PAINLEVÉ.......................... 25 fr.

Principes de la théorie des fonctions entières de genre infini, par OTTO BLUMENTHAL .... 25 fr.

Leçons sur la théorie de la croissance, professées à la Faculté des Sciences de Paris par E. BOREL, recueillies et rédigées par *A. Denjoy*. 25 fr

Leçons sur les séries de polynomes à une variable complexe, par PAUL MONTEL........... .... 25 fr.

Leçons sur le prolongement analytique, professées au Collège de France, par LUDOVIC ZORETTI........ .... 25 fr.

Leçons sur les équations intégrales et les équations intégro-différentielles, professées à la Faculté des Sciences de Rome en 1910, par VITO VOLTERRA, publiées par *M. Tomassetti* et *F.-S. Zarlatti*....... .... 25 fr.

Leçons sur les singularités des fonctions analytiques, professées à l'Université de Budapest, par PAUL DIENES .... 25 fr.

Leçons sur les fonctions de lignes, professées à la Sorbonne en 1911, par VITO VOLTERRA, recueillies et rédigées par *J. Pérès*......... 25 fr.

Les systèmes d'équations linéaires à une infinité d'inconnues, par FRÉDÉRIC RIESZ .... 25 fr.

Intégrales de Lebesgue. Fonctions d'ensemble. Classes de Baire, par C DE LA VALLÉE POUSSIN........ 25 fr

Leçons sur les méthodes de Sturm dans la théorie des équations différentielles linéaires et leurs développements modernes, professées à la Sorbonne en 1913-1914, par MAXIME BÔCHER, recueillies et rédigées par GASTON JULIA........ 25 fr.

Leçons sur les fonctions monogènes uniformes d'une variable complexe, par EMILE BOREL rédigées par G. JULIA........ 25 fr.

Leçons sur l'approximation des fonctions d'une variable réelle, professées à la Sorbonne par C. DE LA VALLÉE POUSSIN........ 25 fr

Leçons sur les fonctions automorphes Fonctions automorphes de n variables. Fonctions de Poincaré, par GEORGES GIRAUD ... 25 fr.

Méthodes et problèmes de Théorie des Fonctions, par É. BOREL 25 fr

Leçons d'analyse fonctionnelle, professées au Collège de France, par PAUL LÉVY, Professeur à l'École Polytechnique, avec une *Préface* de J HADAMARD, Membre de l'Institut........ 50 fr

Leçons sur les fonctions uniformes à point singulier essentiel isolé, professées au Collège de France par GASTON JULIA, Professeur à la Faculté des Sciences de Paris, rédigées par P. FLAMANT ... 25 fr

L'Analysis Situs et la Géométrie algébrique, par S LEFSCHETZ. 25 fr.

Leçons sur les fonctions quasi-analytiques, professées au Collège de France par T. CARLEMAN, Professeur à l'Université de Stockholm. 40 fr

Leçons sur la Composition et les Fonctions permutables, par VITO VOLTERRA, Membre de l'Institut, Professeur à l'Université de Rome, et JOSEPH PÉRÈS, Professeur à l'Université d'Aix-Marseille ..... 25 fr

Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle, professées à la Sorbonne par SERGE BERNSTEIN, Correspondant de l'Institut, Membre de l'Académie des Sciences d'Ukraine .... 50 fr

Leçons sur les Séries d'interpolation, par N.-E. NÖRLUND, Professeur à l'Université de Copenhague, Membre correspondant de l'Institut de France, rédigées par RENÉ LAGRANGE........ 50 fr

Leçons sur les Familles normales de fonctions analytiques et leurs applications, par P. MONTEL, Professeur à la Faculté des Sciences de Paris, recueillies et rédigées par J. BARBOTTE........ 50 fr.

Leçons sur les nombres transfinis, par W. SIERPINSKI, Membre de l'Académie polonaise des Sciences et des Lettres, Professeur à l'Université de Varsovie........ 40 fr

RECHERCHES THÉORIQUES MODERNES

SUR

LA THÉORIE DES PROBABILITÉS

PREMIER LIVRE

Généralités sur les Probabilités. Variables aléatoires

OUVRAGES DU MÊME AUTEUR

Mathématiques appliquées

L'équation de Fredholm et ses applications à la physique mathématique, en collaboration avec M. H. B. Heywood, avec une Préface et une Note de M. J. Hadamard. In-8, 165 pages (*epuisé*). Hermann, Paris, 1912.

Le calcul des probabilités à la portée de tous, en collaboration avec M. Halbwachs. In-16, 300 pages, 18 exemples, 18 figures et 24 tableaux numériques. Dunod, Paris, 1924.

Nomographie (pratique et construction des abaques), en collaboration avec M. H. Roullet (*Collection Colin*, n° 63, section mathématique). In-16, 208 pages, 79 figures. Colin, Paris, 1928.

Représentation des lois empiriques par des formules approchees, à l'usage des chimistes, des physiciens, des ingénieurs et des statisticiens, en collaboration avec M. R. Romann. In-8, VII-302 pages. Eyrolles, Paris, 1930.

Recherches modernes sur la théorie des probabilités (fascicule 3 du tome I du *Traité du Calcul des Probabilités*, par E. Borel et divers auteurs). Premier Livre : *Généralités sur les Probabilités. Variables aléatoires*. In-8, XVI-308 pages. Gauthier-Villars, Paris, 1936.

Mathématiques pures

Développements en séries (*Recherches contemporaines sur la théorie des fonctions de variables reelles*, rédigées sous la direction de M. E. Borel, 3e partie, 1912); *Encyclopédie des Sciences mathématiques pures et appliquees*, édition française, t. II, vol. 1, p. 210-241.

Les espaces abstraits et leur théorie considérée comme introduction à l'analyse générale (*Collection de monographies sur la théorie des fonctions*, dirigée par M. E. Borel). In-8 de XI-296 pages. Gauthier-Villars, Paris, 1928.

L'arithmétique de l'infini, *premier des Exposés d'Analyse générale*, publiés sous la direction de M. Maurice Fréchet. In-8, 57 pages. Hermann, Paris, 1933.

Leçons sur les séries trigonométriques. In-4° dactylographie de 62 pages (Collection : *Les Cours de la Sorbonne*), Tournier et Constant, Paris, 1935.

Théorie élémentaire des équations différentielles. In-4° dactylographie de 56 pages. Tournier et Constant, Paris, 1936.

Notice sur les travaux scientifiques de M. Maurice Fréchet. In-4° de 104 pages, Hermann, Paris, 1933.

TRAITÉ du CALCUL des PROBABILITÉS et de ses APPLICATIONS

Par Émile BOREL

AVEC LA COLLABORATION DE

C.-V.-L. CHARLIER, R. DELTHEIL, P. DUBREIL, M. FRÉCHET, H. GALBRUN,
J. HAAG, R. LAGRANGE, F. PERRIN, Ch. RISSER, P. TRAYNARD

TOME I

LES PRINCIPES DE LA THÉORIE DES PROBABILITÉS

FASCICULE III

RECHERCHES THÉORIQUES MODERNES SUR LA THÉORIE DES PROBABILITÉS

PREMIER LIVRE

Généralités sur les Probabilités. Variables aléatoires

Par Maurice FRÉCHET

PROFESSEUR
DE CALCUL DIFFÉRENTIEL ET INTÉGRAL
A LA FACULTÉ DES SCIENCES DE PARIS

Avec une Note de Paul LÉVY

PARIS

GAUTHIER-VILLARS, ÉDITEUR

LIBRAIRE DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE

55, Quai des Grands-Augustins, 55

1937

PRÉFACE.

M. Émile Borel avait d'abord l'intention de rédiger lui-même ce fascicule. Mais ses multiples occupations lui ont fait craindre de ne pouvoir réaliser son dessein et il m'a fait l'honneur de me proposer de le remplacer.

L'ensemble des travaux modernes sur le Calcul des Probabilités est considérable, même quand on ne tient compte que des plus importants. J'ai donc dû me résigner à n'en résumer d'abord qu'une partie pour ne pas trop retarder la publication de ce fascicule. Comme j'avais eu la possibilité d'exposer et de compléter certaines de ces recherches modernes dans mes cours à l'Institut Henri Poincaré, dans la période 1928-1933, il m'a paru que le mieux était de consacrer ce fascicule aux mêmes recherches, sur lesquelles j'avais eu le plus à réfléchir. Mais, pour en faciliter la rédaction, on a divisé ce fascicule en deux parties : le Premier et le Second Livre qui seront imprimés successivement. Le Premier Livre est principalement consacré à la théorie des variables aléatoires qui est venue simplifier et renouveler la théorie classique des probabilités. Le Second Livre traite de la méthode des fonctions arbitraires et de la théorie des probabilités en chaîne, qui ont leurs origines dans les conceptions profondes de Poincaré et de Markoff et qui fournissent à la Physique mathématique de précieux moyens

d'investigation. Comme ce Traité contient des fascicules concernant spécialement les applications des Probabilités à la Physique mathématique, je me suis limité à l'exposé des résultats mathématiques et je n'ai cité d'applications que dans la mesure où elles pouvaient illustrer la théorie. Enfin, le fascicule de ce Traité prévu sous le nom de « Compléments divers » se prêtera à l'exposé des travaux qui n'ont pu être résumés dans ce fascicule et dont certains sont d'une très grande importance.

Ces trois ouvrages ont, avant tout, pour but de rassembler les résultats obtenus dans ces dernières années en Calcul des Probabilités par toute une génération de chercheurs. Ils auraient paru beaucoup plus tôt si mon intervention s'était bornée à replacer ces résultats dans un ordre systématique. Mais j'ai eu l'imprudence de trop m'intéresser à ces recherches et de m'y engager moi-même. Au moment où j'ai accepté de rédiger ce fascicule, je venais à peine de terminer mon ouvrage *Les Espaces abstraits* et mes premières contributions au Calcul des Probabilités étaient tout naturellement en connection directe avec l'ordre d'idées qui règne en Analyse générale (¹), comme on le verra aux Sections III, IV et V du Chapitre V du Premier Livre. Mais la lecture de certaines publications à résumer en vue de ce fascicule, m'a conduit à procéder à une révision de ces travaux d'une toute autre nature, pour en renforcer la rigueur d'une part, pour les compléter ou les prolonger d'autre part. Alors qu'une grande partie de mes recherches en Analyse générale a été consacrée à une œuvre d'exploration et de construction, il

(¹) Pour la signification de ce mot, voir la série d'*Exposés sur l'Analyse générale*, publiée sous ma direction par la librairie Hermann et particulièrement la préface du premier de ces Exposés : *L'Arithmétique de l'infini*.

s'agissait donc, cette fois, d'exploiter un domaine plutôt que de le découvrir. On verra dans le Second Livre que j'ai pu apporter des réponses complètes à certaines questions, posées mais non résolues, de la Théorie des probabilités en chaîne, en employant comme mes prédécesseurs les procédés de l'Analyse la plus classique.

Mais les applications du Calcul des Probabilités commandent aussi tout un ensemble de recherches à effectuer avec le concours de l'Analyse générale. C'est seulement parce qu'on n'avait encore les moyens, ni de résoudre les problèmes correspondants, ni même de poser clairement ceux-ci, qu'on ne les avait pas encore abordés jusqu'à une époque toute récente. C'est là une grande œuvre collective à entreprendre à laquelle je me propose d'apporter une collaboration active dès la fin de mes travaux en cours.

On ne pourra en recueillir tous les fruits que dans une nouvelle édition, que je souhaite aussi rapprochée que possible, de l'ensemble des volumes de ce Traité . le premier à pouvoir être comparé — au moins par l'étendue des matières étudiées — à l'inépuisable *Traité analytique des probabilités* de Laplace.

MAURICE FRECHET

27 octobre 1935

PREMIER LIVRE

Généralités sur les Probabilités. Variables aléatoires

SOMMAIRE DU PREMIER LIVRE (1)

Pages

AVERTISSEMENT . . . xv

PREMIÈRE PARTIE

GÉNÉRALITÉS SUR LES PROBABILITÉS

CHAPITRE I. — *La notion de probabilité* . . . 1

CHAPITRE II. — *Diverses extensions du principe des probabilités totales* . . . 11

SECONDE PARTIE

LES VARIABLES ALÉATOIRES

CHAPITRE III. — *Valeurs moyennes des variables aléatoires.*

Section I. Introduction . . . 29

Section II. Valeurs moyennes . . . 35

Section III. Épreuves répétées . . . 84

Section IV. Historique . . . 105

CHAPITRE IV. — *L'inégalité de Bienaymé et ses généralisations* . . . 111

CHAPITRE V. — *Les divers modes de convergence d'une suite de variables aléatoires.*

Section I. Introduction . . . 158

Section II. Convergence en probabilité . . . 164

Section III. Premier espace des variables aléatoires . . . 191

Section IV. Convergences de diverses natures et espaces correspondants . . . 200

Section V. Convergence presque certaine . . . 215

Section VI. Suites asymptotes . . . 264

Supplément mathématique. — *Rappel des propriétés des fonctions monotones* . . . 270

Note A. *Une propriété nouvelle de la seconde loi de Laplace* . . . 277

Note B, par M. Paul Lévy. — *Distance de deux variables aléatoires et distance de deux lois de probabilité* . . . 286

Note C. *Additions diverses* . . . 293

LISTE BIBLIOGRAPHIQUE . . . 297

TABLE DES MATIÈRES . . . 303

(1) On trouvera à la fin du Premier Livre (p. 303) une table des matières plus détaillée.

AVERTISSEMENT.

Le sommaire qui précède donne un premier aperçu du contenu de ce Premier Livre (terminé d'ailleurs par une Table complète des matières). On verra que les matières traitées dans les cinq Chapitres sont assez différentes les unes des autres. Ceci ne saurait d'ailleurs étonner dans un ouvrage prétendant résumer un ensemble de travaux isolés parus dans un grand nombre de périodiques sous des signatures diverses. (Pour ne pas interrompre la lecture, les références bibliographiques ont été réduites dans le texte à des numéros se rapportant aux titres complets rassemblés et reportés à la fin de l'ouvrage).

Ce Premier Livre aurait pu être rédigé plus succinctement. L'auteur est entré dans des détails qui paraîtront superflus aux lecteurs les plus avancés. Le contenu de ce livre ayant fait l'objet de plusieurs cours à la Faculté des Sciences de Paris, l'auteur a cru qu'il pourrait y avoir des avantages à conserver ici le mode d'exposition plus explicite qui convient à l'enseignement. La lecture de l'ouvrage en sera ainsi facilitée à un grand nombre de ceux qui utilisent le Calcul des Probabilités sans être des mathématiciens de profession.

J'adresse en terminant des remerciements très vifs à MM. Eyraud, Fortet, Gumbel, Lelong qui ont bien voulu

m'aider dans la tâche aride que constitue la correction des épreuves. J'ai eu en outre le grand privilège de recevoir des avis scientifiques très précieux ou d'utiles renseignements bibliographiques de MM. Cantelli, Cramér, Glivenko, Kolmogorroff, Paul Lévy, de Misès et Slutsky, auxquels j'exprime ici toute ma reconnaissance.

A cette collaboration, il me serait agréable d'adjoindre celle des lecteurs de ce Livre, dont j'accueillerai avec intérêt les observations.

Je remercie enfin la maison Gauthier-Villars, à qui je dois, en particulier, d'avoir pu tenir compte dans ce livre, de quelques travaux portés à ma connaissance tout récemment et même pendant la correction des épreuves.

MAURICE FRECHET

RECHERCHES MODERNES

SUR LE

CALCUL DES PROBABILITÉS

PREMIÈRE PARTIE.

GÉNÉRALITÉS.

CHAPITRE I.

LA NOTION DE PROBABILITÉ

Les diverses définitions de la probabilite. — Différentes définitions ont été données de la probabilité, l'origine de chacune d'elles peut être trouvée dans la question concrete qui a nécessité, pour être formulée de façon précise, l'énoncé d'une telle définition.

Les plus anciens problèmes de probabilité sont ceux qui se sont posés a l'occasion des jeux de hasard. La pratique a conduit ceux-ci à une forme telle qu'on y puisse aisément discerner des cas également possibles. Par exemple, quand on jette un dé, on ne voit pas de raison pour que le point obtenu soit 1, 2, 3, 4, 5 ou 6.

La définition classique. — La premiere définition qui a été donnée de la probabilité repose donc sur la double hypothèse : I, que les diverses modalités du résultat de chaque épreuve peuvent être réparties en un nombre fixé N de cas *également* possibles qui s'excluent mutuellement et : II, que l'événement E envisagé se produit nécessairement dans R de ces cas dits cas favorables (à E) et ne peut se produire dans les autres, dits cas défavorables (à E).

Alors la probabilité de l'événement est égale par définition à $\frac{R}{N}$. C'est la définition qui a été rendue classique par Laplace. Il y a deux objections principales à cette définition : 1° si elle réduit la définition de la *valeur* de la probabilité à celle de l'*égalité* de deux probabilités, elle ne définit pas cette égalité, 2° elle ne s'applique pas lorsque le nombre des cas également possibles est infini. On peut compléter cette définition de façon à diminuer beaucoup la portée de ces deux objections. C'est ce que nous avons fait dans un cours à la Faculté des Sciences de Paris en 1928-29, et c'est ce que nous montrerons, entre autres, dans une petite brochure en préparation sous le titre *Les diverses définitions de la probabilité*.

Les définitions basées sur la fréquence. — La seconde catégorie de problèmes d'ordre pratique où s'est introduite la notion de probabilité est celle des questions qui se sont posées d'abord dans la théorie des assurances et plus tard en statistique. Dans ces problèmes, ainsi que dans ceux où interviennent les probabilités dites continues ou géométriques, il n'est généralement pas possible de répartir *toutes* les modalités possibles du résultat d'une épreuve en un nombre fini de cas *également* possibles, chacun favorable ou défavorable ([1]). La pratique a conduit à considérer comme valeur empirique de la probabilité d'un événement fortuit sa fréquence dans une longue série d'épreuves. Deux tendances assez différentes ont prévalu quand il s'est agi d'en déduire une définition de la probabilité.

Les collectifs. — Dans l'une, on a fait intervenir individuellement des suites illimitées d'épreuves $e_1, e_2, \dots, e_n, \dots$, c'est ce qu'ont fait autrefois un certain nombre de « théoriciens empiristes » dont le plus notable est Venn. C'est ce que font maintenant plusieurs auteurs qui, à la suite de M. von Mises (4) ([2]), ont voulu fonder une théorie axiomatique de la probabilité en mettant à la base de celle-ci la considération des suites ci-dessus sous le nom nouveau de *collectifs*.

Mais au lieu d'admettre la considération de suites quelconques

---

([1]) Autrement dit, la probabilité n'est pas nécessairement un nombre fractionnaire.

([2]) *Voir*, pour les références complètes, la liste bibliographique placée à la fin du volume, p. 297.

d'épreuves, ils ne désignent sous le nom de collectifs que les suites d'épreuves satisfaisant à certaines conditions dont la plus significative est la suivante. On suppose que la fréquence $\frac{r}{n}$ avec laquelle l'événement fortuit E considéré se présente (r fois dans n épreuves consécutives) a une limite déterminée (qui sera appelée la valeur de la probabilité de E dans le collectif considéré) quand n croît. M. von Mises suppose, en outre, que la suite de ces fréquences jouit d'une certaine irrégularité. La probabilité est donc définie dans sa théorie comme « la limite de la fréquence, cette limite restant invariable — à raison de l'« irrégularité » — quand on supprime une partie des éléments par un choix effectué d'une certaine façon ».

La tentative de M. von Mises est intéressante et l'on doit voir avec sympathie l'effort qu'il a fait pour harmoniser la théorie et la pratique. Toutefois, au dire même de M. von Mises (5, p. 228), on ne doit pas considérer sa théorie comme ayant acquis une forme définitive. D'importantes objections lui ont été faites de divers côtés. MM. Khintchine (4), Cantelli (9) et Copeland (1) ont indiqué qu'elle conduit à des contradictions logiques. Sans entrer dans le détail nous nous contenterons de signaler qu'entre autres, la difficulté suivante ne nous a pas permis jusqu'ici d'adopter la théorie de M. von Mises. Considérons par exemple, le jeu de dés. On peut étudier successivement l'événement E consistant à obtenir indéfiniment le point zéro au moyen d'un dé portant sur chaque face 1, 2, ..., ou 6 points et l'événement E' consistant à obtenir une suite déterminée d'avance, de points dans une série illimitée d'épreuves. Alors que le premier événement est logiquement impossible, le second *quoique possible logiquement* ne se produira pas en pratique parce qu'il a une probabilité nulle.

Le premier événement entrera en considération comme logiquement impossible dans la théorie de M. von Mises de même que dans les autres théories. Le second, logiquement possible dans les autres théories, ne pourra même pas être considéré dans celle de M. von Mises, lorsque la série de points est telle que, par exemple, la fréquence $\frac{r}{n}$ du point un dans n épreuves ne tend pas vers une limite déterminée [1] quand n croît. Nous ne voyons aucun motif pour accepter cette exclusion.

---

[1] C'est, par exemple, ce qui aurait lieu si le point 1 se présentait seulemen

Signalons cependant la réponse que donne M. Copeland à cette objection : il n'y a pas d'inconvénient, dit-il, à exclure ce cas, parce qu'il est impossible de concevoir la réalisation *effective* d'une infinité d'épreuves. D'autre part, M. Copeland et M. Wald (1) ont annoncé que des modifications convenables peuvent restituer à la théorie de M. von Mises une forme exempte de contradictions. [*Voir* aussi, Kamke (1), Ville (1), (2)]

Avant de porter un jugement définitif, nous attendrons donc le résultat de ces modifications ou des modifications qui pourront être présentées ultérieurement à une théorie dont l'auteur a apporté par ailleurs d'importantes contributions au calcul des probabilités

La définition statistique. — La seconde tendance que l'on peut observer dans les points de vue inspirés par les problèmes de la statistique est celle qui consiste à distinguer parmi les événements, des événements dits fortuits obéissant à la loi du hasard. [*Voir*, par exemple, Cournot (1)].

Dans un ouvrage destiné aux débutants (F. et H.) [1], M. Halbwachs et moi-même ont cru à la fois légitime, et plus facile pour ces débutants, de nous placer à un point de vue analogue et de partir d'une définition empirique de la probabilité. On peut résumer notre définition comme suit

Le hasard. — Nous admettons ici le hasard comme une notion familière et nous supposons qu'on sait distinguer parmi les événements, ceux qui sont fortuits.

Dans un groupe déterminé de n épreuves, un événement fortuit s'est produit r fois : nous appellerons r la *répétition* de cet événement et $\frac{r}{n}$ sa *fréquence* dans le groupe de n épreuves. Nous allons maintenant admettre, non comme un postulat, mais comme une vérité que nous fait apparaître notre observation quotidienne, la loi expérimentale du hasard.

---

aux épreuves de rangs 1, 17 à 32, . . ., $2^{r}+1$ à 2^{r+1}, . . On aurait alors $f_n = \frac{1}{3}\,\frac{4^r-1}{4^r}$ pour $n = 2^{2r}$, $f_n = \frac{2}{3}\,\frac{4^{r+1}-1}{4^{r+1}}$ pour $n = 2^{2r+1}$, de sorte que f_n ne pourrait avoir une limite unique

(1) Dans la suite, l'abréviation « F et H. » désignera l'ouvrage élémentaire intitulé *Le Calcul des Probabilités à la portée de tous* par Fréchet et Halbwachs, chez Dunod, Paris, 1924

Loi expérimentale du hasard. — On constate que, dans une catégorie déterminée d'épreuves, lorsqu'on calcule les fréquences d'un événement fortuit dans différents groupes d'épreuves, ces fréquences sont, pratiquement, peu différentes quand chacun de ces groupes est constitué de nombreuses épreuves En d'autres termes, ces fréquences « se massent » autour d'une même valeur qui est d'autant mieux déterminée que ces groupes contiennent plus d'épreuves.

On peut alors énoncer la loi du hasard sous la forme suivante :

Les fréquences d'un événement fortuit E *dans des groupes comprenant chacun de nombreuses épreuves appartenant toutes à une même catégorie* C *sont les valeurs expérimentales d'une même constante physique déterminée par la nature de l'événement* E *et par la catégorie* C

Définition empirique de la probabilité. — C'est cette constante physique que nous appellerons *la probabilité de l'événement fortuit* E, *dans la catégorie d'épreuves* C.

Un tel nombre est déterminé par E et C au même sens qu'on peut considérer comme bien déterminée la mesure de la longueur d'une barre.

On peut ajouter à l'énoncé de la loi du hasard que la précision de la mesure de la probabilité p de E par sa fréquence f_n dans un groupe de n épreuves est pratiquement d'autant plus grande (nous ne disons pas que l'erreur ainsi commise est d'autant plus petite) que n est plus grand. Mais il doit être bien entendu qu'on doit distinguer cette observation de celle qui consisterait à écrire

$$p = \underset{n \to \infty}{\text{limite}}\, f_n$$

égalité qui, prise strictement, est fausse.

De même qu'on s'attend seulement, en effectuant deux mesures d'une même longueur, à obtenir deux nombres voisins, de même, on « s'attendra » seulement en effectuant deux séries illimitées d'épreuves à trouver la même limite pour la fréquence d'un même événement fortuit.

Remarque. — Il est clair que toute fréquence, par sa définition même, est un nombre compris entre zéro et l'unité, que la fréquence d'un événement certain est égale à l'unité et celle d'un événement impossible est nulle. Il en sera donc de même de la probabilité.

Axiomatisation. — Enfin, une dernière manière d'éviter les objections d'ordre pratique consiste dans une rigoureuse axiomatisation du calcul des probabilités. Les théorèmes des probabilités totales et composées qui se trouvent démontrés à partir de l'une quelconque des trois définitions envisagées ci-dessus, sont dans cette dernière théorie, admis sans explication, à titre de postulats. A chaque événement fortuit, on suppose attaché un nombre, appelé probabilité de cet événement, qui est seulement soumis aux postulats suivants : 1° il est $\geqq 0$ et $\leqq 1$; 2° il vérifie les théorèmes des probabilités totales et composées. C'est la position qui a été prise implicitement par M. Borel dans le premier Fascicule de ce Traité. C'est à peu près celle qui a été détaillée explicitement par MM. Cantelli (9, § 8) et Kolmogoroff (4).

On doit noter une tendance, dans ces dernières années, à modeler la théorie des probabilités sur la théorie de la mesure. A vrai dire, ce n'est pas la notion de mesure d'un ensemble, c'est plutôt la notion de *fonction additive* (non négative) d'ensemble, qui est propre, en raison du théorème (ou du postulat) des probabilités totales à représenter une probabilité soit continue, soit discontinue. Il est vrai qu'on peut dans bien des cas, comme l'a fait observer M. Borel, déterminer un changement de variables transformant une fonction additive (non négative) d'ensemble dans la mesure de l'ensemble transformé. Toutefois un des avantages de la considération des fonctions additives d'ensembles, c'est qu'elle permet d'envisager à la fois les probabilités continues et discontinues. Or, pour ces dernières, le changement de variable mentionné ci-dessus serait généralement impossible.

Importance de la catégorie d'épreuves. — Dès nos premières réflexions sur le calcul des probabilités, nous avons été frappé de l'insuffisance de cette expression : la probabilité d'un événement. En principe, il ne suffit pas de préciser l'événement considéré pour déterminer sa probabilité. Pourtant il n'y a pas généralement d'inconvénient à se contenter de mentionner l'événement parce que la catégorie d'épreuves où l'on calcule sa probabilité est sous-entendue d'une façon assez claire. Par exemple, si l'on parle de la probabilité d'amener 3 au jeu de dés, on sait bien de quelles épreuves il s'agit. De même, si l'on calcule la probabilité de tirer un as de cœur, il

s'agit de le tirer d'un jeu de cartes complet retourné et bien battu. Toutefois déjà, dans ce cas, il y a lieu d'observer que, dans certains jeux, on emploie des paquets de 32 cartes, dans d'autres cas des paquets de 52 cartes. Il faudra donc préciser.

Mais, il y a des cas encore plus douteux. Bien des difficultés qui se sont présentées en calcul des probabilités sont dues à ce qu'on n'a pas précisé suffisamment la catégorie d'épreuves où la probabilité d'un événement était considérée. C'est, par exemple, le cas du paradoxe de Bertrand, il s'agit de déterminer la probabilité P pour qu'une corde d'un cercle donné soit plus petite que le côté d'un triangle équilatéral inscrit. Trois raisonnements différents conduisent à trois valeurs différentes de P. Cela tient à ce que l'on n'a pas précisé suffisamment la réalisation matérielle du choix fortuit de la corde. On peut lancer une aiguille de longueur plus grande que le diametre du cercle et ne compter pour épreuves que celles où l'aiguille aura deux points en contact avec le cercle. Ces deux points détermineront la corde. Ou bien, on fait tourner deux fois une aiguille autour du centre du cercle et l'on joint les deux points d'arrêts de l'extrémité de l'aiguille sur le cercle, etc. A chaque mode de réalisation correspondra une valeur de la probabilité

La considération de la catégorie d'épreuves fait disparaître un autre paradoxe la probabilite d'un événement serait d'autant mieux determinée que nos renseignements sur cet événement seraient moins nombreux Plus nous serions ignorants, plus précise serait notre connaissance de la probabilité.

Il y a là une antithese choquante qui s'évanouit quand on a soin de ne jamais parler de la probabilité d'un événement sans préciser dans quelle catégorie d'épreuves cette probabilité est définie.

Il n'est alors pas nécessaire de faire intervenir notre ignorance pour nous autoriser à élargir une catégorie.

Je veux me former une idée de la probabilité au sens vulgaire qu'a M. Dupont, rentier, de mourir l'année prochaine. Même si je possede beaucoup de renseignements sur lui, sur son âge, sur la couleur de ses cheveux, le nombre de ses dents, ..., je n'aurai aucun intérêt à calculer la probabilité qu'il a de mourir en prenant pour valeur la fréquence de la mort parmi tous les individus possédant à la fois les mêmes caractéristiques. Car en faisant intervenir successivement toutes celles-ci, on réduit de plus en plus le nombre de ces individus

et, par suite, on a de moins en moins le droit de leur appliquer la loi du hasard. Si donc, je me contente, pour estimer la probabilité en question, de prendre la probabilité de mortalité annuelle parmi les rentiers français du même âge (table R. F.), ce n'est pas parce que je n'ai pas sur lui d'autres renseignements que la valeur choisie de la probabilité est meilleure, c'est parce que, parmi mes renseignements, j'ai justement retenu ceux seulement qui devaient la rendre meilleure.

L'importance de la notion de catégorie d'épreuves que nous avons tenu, M. Halbwachs et moi, à mettre en lumière (F. et H., p. 10) a toujours été reconnue, mais en général seulement d'une façon implicite, par les mathématiciens qui se sont occupés de calcul des probabilités

Certains ont expressément insisté sur une idée voisine. M. von Mises écrit (4, p. 141) : « l'expression « probabilité » sans cette adjonction « dans le collectif considéré, » n'existe pas pour nous.... Chaque individu appartenant à des classes diverses, on ne peut jamais parler d'une probabilité de survie pour l'individu lui-même sans se rapporter à un collectif bien défini. » C'est là une observation très importante. Il y a cependant à distinguer ici la population, qui est l'ensemble des individus considérés et la catégorie d'épreuves. Si par exemple, il s'agit de la survie à l'âge de 40 ans. cette catégorie sera l'ensemble des épreuves consistant chacune dans la vie d'un individu appartenant à une population déterminée à partir du quarantième anniversaire et dans l'observation de tout ce qui se passe, en particulier de sa survie ou de son décès, au cours de sa quarantième année.

La notion de catégorie d'épreuves est claire par elle-même. Par exemple, dans le jeu de dés, la catégorie d'épreuves est l'ensemble des expériences individuelles consistant « dans le processus suivant : un dé est mis dans un cornet, on secoue le cornet, on jette le dé sur la table et on lit le chiffre apparaissant en haut ». Cependant, M. von Mises (qui précise dans ces termes entre guillemets un exemple d'« expérience individuelle ») se demande si la notion de catégorie d'épreuves a été suffisamment précisée. Ce qui va sans dire va encore mieux en le disant. Ajoutons donc, d'abord, que dans l'exemple ci-dessus, on aurait eu encore la même catégorie d'épreuves, mais un autre jeu, si, au lieu de noter le point obtenu, on avait relevé, par exemple, les angles des faces verticales du dé, après sa chute, avec une horizontale donnée. Disons, d'une façon générale, qu'une catégorie d'épreuves est l'ensemble des épreuves consistant dans la

réalisation d'un phénomène déterminé et dans l'observation des caractéristiques du résultat de l'épreuve. L'événement correspondra à l'une des modalités possibles de l'une de ces caractéristiques. Par exemple, la catégorie d'épreuves peut être l'ensemble des jets de dé et l'événement consistera dans le fait que l'angle aigu d'une face verticale avec une horizontale donnée est compris entre 20° et 30°. Par exemple, encore, la catégorie d'épreuves peut consister dans l'ensemble des décès des Français nés en 1848, et l'observation des âges respectifs atteints au décès, ou de leurs professions respectives à l'époque de la mort. Les âges, les professions sont deux des caractéristiques des résultats des épreuves.

L'événement dont on cherche la probabilité (dans cette catégorie d'épreuves) sera, par exemple, l'âge de 30 ans dans le premier cas, ou par exemple, la profession de forgeron dans le second cas. La modalité choisie est ici précisée par le nombre 30 ou par le mot forgeron.

Categories complexes — Observons enfin qu'on a souvent à envisager des événements E qui sont des combinaisons déterminées de 2, 3, . ., n, ou même d'une infinité d'événements $E_1, E_2, \ldots$. Par exemple, E consiste en ce que E_1 a lieu mais non E_2; par exemple encore, E consiste dans la réalisation d'une infinité des événements E_n. Quand $E_1, E_2, \ldots$ sont définis sur des categories d'épreuves $C_1, C_2, \ldots$ E sera défini sur une catégorie C d'épreuves dont chacune s'obtient en associant des épreuves de $C_1, C_2, \ldots$. Dans le premier exemple, on se contentera d'associer une épreuve de C_1 et une épreuve de C_2. Si les deux catégories C_1, C_2 ne sont pas indépendantes, l'association ne pourra d'ailleurs avoir lieu de toutes les façons possibles

Nous ajouterons encore quelques mots (p. 16-22), a ces réflexions générales sur la notion de probabilité. Celle-ci sera étudiée plus longuement dans le fascicule intitulé « Conclusions : la portée philosophique du Calcul des Probabilités » qui sera le facicule 4 du tome IV du présent Traité. D'autre part, nous exposerons plus en détail nos conclusions personnelles dans un Exposé mentionné déjà plus haut (p. 2) sous le titre « Les diverses définitions de la probabilité ».

Répétition, fréquence. — Disons un mot seulement pour avertir que nous avons cru devoir utiliser le mot répétition pour éviter, dans

ce fascicule, une confusion qui se produit souvent dans le sens du mot fréquence. Si par exemple, on a jeté 100 fois une pièce de monnaie et obtenu 48 fois pile, nous dirons que 48 est la *répétition* de l'événement pile dans les 100 épreuves et que $\frac{48}{100}$ est la *fréquence* correspondante.

CHAPITRE II.

DIVERSES EXTENSIONS DU PRINCIPE DES PROBABILITÉS TOTALES.

On sait (I, 1, p. 2) [1] que si un événement fortuit E peut se produire suivant un nombre fini de modalités incompatibles $E_1, E_2, \dots, E_n$, sa probabilité est égale à la somme des probabilités de chacune de ces modalités

$$\text{Pr}\{E\} = \text{Pr}\{E_1\} + \dots + \text{Pr}\{E_n\} \tag{1}$$

On peut généraliser ce principe de plusieurs manières.

Cas d'un nombre fini d'evénements compatibles. — On peut d'abord se demander quels renseignements sur le premier membre peuvent résulter de la connaissance des termes du second membre quand les modalités $E_1, E_2, \dots, E_n$ cessent d'être incompatibles.

Pour cela, appelons d'abord q_1, q_2 les probabilités respectives de deux événements (compatibles ou non) E_1 et E_2, q_{12} la probabilité du concours de E_1 et de E_2, Q_{12} la probabilité de la réalisation de l'une au moins de ces modalités, Q'_{12} la probabilité qu'une seule se realise, q'_1 la probabilité que E_1 se réalise et non E_2, et de même q'_2.

On a évidemment

$$Q_{12} = q_1 + q'_2, \qquad q_1 = q_{12} + q'_1, \qquad q_2 = q_{12} + q'_2, \qquad Q'_{12} = q'_1 + q'_2.$$

D'où une première série de formules importantes :

$$q'_1 = q_1 - q_{12}, \qquad Q_{12} = q_1 + q_2 - q_{12}, \qquad Q'_{12} = q_1 + q_2 - 2q_{12}. \tag{2}$$

[1] Dans la suite, une réference telle que « I, 1, p. 2 » renverra à la page 2 du fascicule 1 du tome I de cet ouvrage.

On passera facilement au cas de trois événements F_1, F_2, F_3. Il suffira d'appliquer les formules précédentes en prenant pour E_1 l'événement F_1 et pour E_2 l'événement consistant dans la réalisation de F_2 ou F_3. On passera de même par récurrence au cas général d'un nombre fini quelconque d'événements fortuits H_1, H_2, ..., H_n compatibles ou non. Appelons p_1 la probabilité de H_1, ... p_n celle de H_n, p'_i la probabilité que H_i soit le seul réalisé de ces n événements, $p_{ih \ldots s}$ la probabilité du concours des événements H_i, H_h, ..., H_s, P la probabilité que l'un au moins des n événements H_1, ..., H_n se réalise, P′ la probabilité qu'un seul d'entre eux se réalise. La méthode de récurrence ci-dessus mentionnée conduira aux trois formules fondamentales suivantes :

$$(3) \qquad p'_i = p_i - \sum_h p_{i,h} + \sum_{hk} p_{i,h,k} - \ldots + (-1)^{n-1} p_{1,2,\ldots,n},$$

$$(4) \qquad P = \sum_i p_i - \sum_{i,h} p_{i,h} + \sum_{i,h,k} p_{i,h,k} - \ldots + (-1)^{n-1} p_{1,2,\ldots,n},$$

$$(5) \qquad P' = \sum_i p_i - 2\sum_{i,h} p_{i,h} + 3\sum_{i,h,k} p_{i,h,k} - \ldots + (-1)^{n-1} n p_{1,2,\ldots,n}$$

Parmi ces trois formules classiques, la seconde due à Poincaré (*Calcul des Probabilités*, p. 60) peut être appliquée à un cas particulier pour fournir ainsi (Fréchet, 177) une formule démontrée directement par M. Charles Jordan (1). Considérons des événements *incompatibles* A_1, A_2, ... et appliquons la formule de Poincaré en appelant H_j l'événement consistant en ce qu'au cours d'un nombre fixe r, d'épreuves indépendantes, l'événement A_j ne s'est jamais présenté. Si la probabilité ϖ_j de A_j est constante, on aura

$$p_i = (1 - \varpi_i)^r, \qquad \ldots, \qquad p_{i,j,k} = (1 - \varpi_i - \varpi_j - \varpi_k)^r,$$

D'où, en posant $\varpi = 1 - P$, la formule de M. Jordan

$$(4\,bis) \qquad \begin{aligned} \varpi = 1 &- \Sigma(1-\varpi_i)^r + \Sigma(1-\varpi_i-\varpi_j)^r \\ &- \Sigma(1-\varpi_i-\varpi_j-\varpi_k)^r + \ldots \\ &+ (-1)^n(1-\varpi_1-\varpi_2-\ldots-\varpi_n)^r \end{aligned}$$

qui donne, connaissant les probabilités ϖ_j des A_j, la probabilité ϖ pour que chacun des événements incompatibles A_1, ..., A_n se pro-

duise au moins une fois au cours des n épreuves. Revenons maintenant à la formule initiale de Poincaré.

Dans le cas des événements H incompatibles, les $p_{i,h}$ sont nuls et la seconde égalité se réduit à $P = \sum_i p_i$. Peut-on, connaissant seulement les p_i, dire quelque chose sur P dans le cas général ?

En reprenant les notations ci-dessus, on déduit des relations

$$q_1 + q_2 = q_{12} + Q_{12} \quad \text{et} \quad 0 \leqq q_{12} \leqq 1,$$

les relations

$$(6) \qquad q_1 + q_2 - 1 \leqq Q_{12} \leqq q_1 + q_2$$

En employant la méthode de récurrence ci-dessus indiquée, on en déduit facilement les relations

$$(7) \qquad p_1 + \ldots + p_n - (n-1) \leqq P \leqq p_1 + \ldots + p_n$$

On peut aussi les obtenir directement. La première inégalité s'obtient en ajoutant les inégalités évidentes

$$p_1 \leqq P, \quad p_2 \leqq 1, \quad \ldots, \quad p_n \leqq 1$$

La seconde s'obtient en remarquant que P est égal à la somme des probabilités p_1, θ_2, θ_3, ... θ_n des événements incompatibles H_1, H_2 sans H_1, H_3 sans H_2 ni H_1, .., H_n seul, et en combinant les relations

$$P = p_1 + \theta_2 + \theta_3 + \ldots + \theta_n, \quad \theta_2 \leqq p_2, \quad \theta_3 \leqq p_3, \quad \ldots \quad \theta_n \leqq p_n.$$

D'ailleurs, on peut renforcer la première des deux inégalités (7) en question. On a, en effet, évidemment

$$p_1 + \ldots + p_n - (n-1) \leqq p_i \leqq P \qquad \text{pour } i = 1, 2, \ldots, n$$

On a donc (Fréchet, 177) finalement le système suivant d'inégalités pour estimer la probabilité P que se produise l'un au moins des événements H_1, H_2, ... H_n, de probabilités p_1, p_2, ..., p_n

$$(8) \qquad \varpi(p_1, p_2, \ldots, p_n) \leqq P \leqq \Pi(p_1, p_2, \ldots, p_n),$$

en appelant $\varpi(p_1, p_2, \ldots, p_n)$ la plus grande des probabilités p_1, ..., p_n et $\Pi(p_1, p_2, \ldots, p_n)$ le plus petit des deux nombres 1 et $p_1 + \ldots + p_n$.

Il est intéressant de noter *qu'on ne peut obtenir des inégalités*

plus précises. En d'autres termes si $f(p_1, p_2, \ldots, p_n)$ et $F(p_1, p_2, \ldots, p_n)$ sont deux fonctions de $p_1, \ldots, p_n$, indépendantes des événements $H_1, \ldots, H_n$ et telles que l'on ait

$$f(p_1, p_2, \ldots, p_n) \leqq P \leqq F(p_1, p_2, \ldots, p_n)$$

pour tout systeme d'événements fortuits $H_1, H_2, \ldots, H_n$ de probabilités $p_1, \ldots, p_n$, on a nécessairement

$$(9) \quad f(p_1, \ldots, p_n) \leqq \varpi(p_1, \ldots, p_n) \leqq \Pi(p_1, \ldots p_n) \leqq F(p_1, \ldots, p_n).$$

Prenons au hasard, les nombres $p_1, \ldots, p_n$ compris entre 0 et 1 et soit, par exemple, p_k, le plus grand de ces nombres On pourra toujours définir un événement H_k de probabilité p_k, puis définir des événements $H_1, \ldots, H_{k-1}, H_{k+1}, \ldots, H_n$ qui n'aient lieu que si H_k a lieu et qui soient de probabilités $p_1, \ldots, p_{k-1}, p_{k+1}, \ldots, p_n$. Pour ces événements, on aura

$$f(p_1, \ldots, p_n) \leqq P = p_k = \varpi(p_1, \ldots, p_n),$$

et la première des inégalités (9) est établie.

Pour établir la seconde, nous distinguerons deux cas :

Si $p_1 + p_2 + \ldots + p_n \leqq 1$, on peut facilement définir des événements fortuits incompatibles de probabilités respectives $p_1, \ldots, p_n$. Pour ces événements, on aura

$$\pi(p_1, \ldots, p_n) = p_1 + \ldots + p_n = P \leqq F(p_1, \ldots, p_n).$$

Si $p_1 + \ldots + p_n > 1$, il existe un entier $k < n$, tel que

$$p_1 + \ldots + p_k \leqq 1 < p_1 + \ldots + p_{k+1}$$

On pourra définir des événements fortuits incompatibles $H_1, \ldots H_k$ de probabilités respectives $p_1, \ldots, p_k$, un événement H'_k incompatible avec les précédents et de probabilité $1 - (p_1 + \ldots + p_k) < p_{k+1}$. On désignera par H_{k+1} l'événement consistant en H'_k ou H''_k, H''_k étant un événement quelconque choisi de façon à ce qu'il soit incompatible avec H'_k et de probabilité $p_{k+1} + p_1 + \ldots + p_k - 1 > 0$. {Ces deux conditions sont compatibles puisque

$$\begin{aligned}\text{Pr.}\{H'_k\} + \text{Pr.}\{H''_k\} &= [1 - (p_1 + \ldots + p_k)] \\ &+ [p_{k+1} + p_1 + \ldots + p_k - 1] = p_{k+1} \leqq 1.\}\end{aligned}$$

Appelons enfin $H_{k+2}, \ldots, H_n$ des événements fortuits quelconques de probabilités $p_{k+2}, \ldots, p_n$. Alors la probabilité P sera égale a 1, et l'on aura encore

$$\pi(p_1, \ldots, p_n) \leqq P \leqq F(p_1, \ldots, p_n).$$

Ainsi, les inégalités (9) sont bien établies dans tous les cas.

On peut obtenir des résultats analogues pour la probabilité $p_{1,2,\ldots,n}$ du concours de $H_1 \ldots H_n$. Il suffit pour cela de partir des inégalités qu'on vient d'établir :

$$(10) \qquad \text{Pr.}\{H_i\} \leqq \text{Pr.}\{H_1 \text{ ou } H_2 \ldots \text{ ou } H_n\} \leqq \text{Pr.}\{H_1\} + \text{Pr.}\{H_2\} + \ldots + \text{Pr.}\{H_n\} \quad (i = 1, 2, \ldots, n)$$

et d'y remplacer les événements H par leurs événements contraires. On obtient

$$1 - p_i \leqq 1 - p_{1,2,\ldots,n} \leqq n - (p_1 + \ldots + p_n)$$

ou

$$(11) \qquad p_1 + p_2 + \ldots + p_n - (n-1) \leqq p_{1,2,\ldots,n} \leqq p_i \qquad (i = 1, 2, \ldots, n)$$

En partant de (8), on serait arrivé au système plus précis

$$(12) \qquad \theta(p_1, \ldots, p_n) \leqq p_{1,2,\ldots,n} \leqq \Theta(p_1, \ldots, p_n)$$

où $\theta(p_1, \ldots, p_n)$ est le plus grand des deux nombres

$$p_1 + \ldots + p_n - (n-1)$$

et zéro et où $\Theta(p_1, \ldots, p)$ est le plus petit des nombres $p_1, \ldots, p_n$. La démonstration concernant (8) montre que le système d'inégalités (12) est le plus précis qu'on puisse donner pour estimer $p_{1,2,\ldots,n}$ connaissant seulement $p_1, \ldots, p_n$ au moyen de bornes indépendantes des événements $H_1, \ldots, H_n$.

Observons que la première inégalité (11) peut s'écrire

$$p_{1,2,\ldots,n} \geqq 1 - q_1 - q_2 \ldots - q_n,$$

si l'on pose

$$q_1 = 1 - p_1, \qquad q_2 = 1 - p_2, \qquad \ldots, \qquad q_n = 1 - p_n \quad [1].$$

---

[1] On sait que si $q_1, q_2, \ldots$ sont des nombres compris entre 0 et 1, on a

$$1 - q_1 - q_2 - \ldots - q_n \leqq (1 - q_1) \ldots (1 - q_n).$$

On a donc

$$1 - q_1 - \ldots - q_n \leqq p_1 \ldots p_n \leqq p_i \qquad (i = 1, \ldots, n).$$

Les deux nombres $p_{1,2,\ldots,n}$ et $p_1 p_2 \ldots p_n$ sont donc tous deux toujours compris entre

C'est l'inégalité due à Boole qu'on peut aussi écrire

$$(13)\qquad \text{Pr.}\{H_1 \text{ et } H_2 \ldots \text{ et } H_n\} \geqq 1 - \text{Pr}\{C(H_1)\} - \ldots - \text{Pr}\{C(H_n)\},$$

en désignant par $C(H_i)$ l'évenement contraire à H_i.

Cette inégalité donne des indications souvent utiles en Statistique (F. et H., p. 52). Nous utiliserons fréquemment les inegalités (8) et (12) dans le Chapitre V traitant des différents modes de convergence.

Cas d'une suite infinie d'événements. — Voyons ce que devient le principe des probabilités totales dans le cas où l'on envisage une suite infinie $E_1, E_2, \ldots, E_n, \ldots$ de modalités du même événement E. Et, pour commencer, supposons les modalités $E_1, E_2, \ldots$ incompatibles.

Si les événements $E_1, E_2, \ldots, E_n \ldots$ ont des probabilités determinées $p_1, p_2, \ldots, p_n, \ldots$, l'événement ($E_1$ ou $E_2 \ldots$ ou E_n) aura, d'après le principe des probabilités totales, la probabilité

$$p_1 + \ldots + p_n$$

On a donc

$$p_1 + p_2 + \ldots + p_n \leqq 1,$$

quel que soit n. Par suite, la série

$$p_1 + p_2 + \ldots + p_n + \ldots$$

à termes $\geqq 0$, est nécessairement convergente si les événements E_1, $E_2, \ldots$ sont incompatibles, et sa somme est $\leqq 1$. Si, en outre, E a une probabilité déterminée P, on a, en appelant r_n la probabilité pour que l'un, au moins, des événements $E_{n+1}, E_{n+2}, \ldots$ se réalise :

$$P = p_1 + p_2 + \ldots + p_n + r_n.$$

D'où

$$P \geqq p_1 + p_2 + \ldots + p_n,$$

quel que soit n, et par suite

$$P \geqq p_1 + p_2 + \ldots + p_n + \ldots$$

Ainsi, du principe des probabilités totales, énoncé pour un nombre fini d'événements, on peut déduire que, quelle que soit la suite infinie

---

$\theta(p_1, \ldots, p_n)$ et $\Theta(p_1, \ldots, p_n)$. Quand les événements $H_1, \ldots, H_n$ sont indépendants, on sait que $p_{1,2,\ldots n} = p_1 \ldots p_n$.

d'événements incompatibles $E_1, E_2, \ldots, E_n, \ldots$, on a

$$(14) \quad \Pr \{ E_1 \text{ ou } E_2 \text{ ou } \ldots \text{ ou } E_n \text{ ou } \ldots \} \geqq \Pr \{ E_1 \} + \Pr \{ E_2 \} + \cdots + \Pr. \{ E_n \} + \cdots.$$

Il n'est pas, *a priori*, impossible d'imaginer une définition de la probabilité conforme aux mêmes principes et pour laquelle on ne pourrait remplacer dans la relation (14), le signe $\geqq$ par le signe $=$.

Considérons en particulier le cas où toutes les probabilités considérées sont relatives aux positions que peut prendre un point X au hasard sur une droite illimitée.

Plaçons-nous dans le cas où la probabilité $\varpi(e)$ pour que le point X tombe sur un ensemble arbitraire e de points 1° satisfait au principe restreint des probabilités totales, c'est-à-dire énoncé seulement pour un nombre fini d'événements incompatibles; 2° est uniforme, c'est-à-dire se conserve dans la superposition.

Alors, si l'on appelle I_n l'ensemble $n \leqq x < n + 1$, on voit que $\varpi(I_n)$ est indépendant de n. Si a est sa valeur, on aura

$$0 \leqq na = \varpi(I_1) + \varpi(I_2) + \cdots + \varpi(I_n) \leqq 1.$$

D'où $a = 0$ puisque a est indépendant de n.

On aura donc

$$\varpi(I_0 + I_1 + I_{-1} + \cdots + I_n + I_{-n} + \cdots)$$
$$= 1 \quad a + a + \cdots = \varpi(I_0) + \varpi(I_1) + \varpi(I_{-1}) + \cdots + \varpi(I_n) + \varpi(I_{-n}) + \cdots$$

Voici donc un exemple simple où le principe *restreint* des probabilités totales n'entraîne pas logiquement le principe *complet* des probabilités totales exprimé par l'égalité

$$(14') \quad \Pr(E_1, \text{ ou } E_2, \text{ ou } \ldots, \text{ ou } E_n, \text{ ou } \ldots) = p(E_1) + p(E_2) + \cdots + p(E_n) + \cdots$$

Passons au cas ou X est assujetti a rester sur un segment borné donné I.

M. Vitali a démontré qu'il n'existe aucune fonction d'ensemble $\varpi(E)$, définie pour tout ensemble de points de I, non négative, telle que $\varpi(E) = \varpi(E_1)$ si E et E_1 sont superposables, telle que $\varpi(I) = 1$ et qui possède la propriété d'additivité complète exprimée par (14′).

Par conséquent, même si l'on se restreint à la probabilité la plus

simple concernant la position d'un point sur un segment, à savoir la probabilité uniforme, c'est-à-dire invariante dans la superposition, il est impossible de définir pour *tout* sous-ensemble de I une telle probabilité qui satisfasse au principe des probabilités totales élargi au cas d'une suite infinie d'événements fortuits incompatibles quelconques.

Plus généralement, MM. Banach (1, 2) et Kuratowski (1) ont montré que l'impossibilité subsiste quand au lieu de supposer la probabilité uniforme, on suppose simplement qu'il s'agit d'une probabilité continue ou plus précisément quand on suppose que la probabilité pour que le point variable X tombe en un point déterminé, est nulle quel que soit ce point.

Tout ceci rend bien clair que du principe des probabilités totales exprimé par la formule (1), on ne peut déduire logiquement l'extension au cas d'un nombre infini d'événements exprimé par (14').

Par contre, *si dans chaque catégorie d'épreuves, l'on n'attribue pas une probabilité déterminée à tout événement fortuit*, il est parfaitement possible de la définir de façon à réaliser cette extension. Par exemple, dans le cas de la répartition uniforme des probabilités relatives à la position d'un point pris au hasard sur un segment de longueur un, on devra attribuer à tout ensemble linéaire mesurable au sens de M. Lebesgue une probabilité égale à sa mesure. Si l'on attribue en outre une certaine probabilité aux ensembles non mesurables, on pourra bien, d'après un autre résultat de M. Banach (1, p. 31), le faire de façon à respecter la formule finie (1), mais on ne pourra d'aucune manière réaliser cette extension pour tous les ensembles non mesurables de façon à vérifier la formule infinie (14') et l'invariance par superposition.

Ceci nous fait comprendre comment on est conduit à certaines antinomies [1] en partant de la formule (14'). Celles-ci s'évanouissent quand on ne perd pas de vue la question de l'existence des probabilités avec lesquelles on opère. D'une part, si l'on utilise la définition empirique de la probabilité (*voir* F. et H.), comme celle-ci suppose que les événements considérés obéissent à la loi du hasard, on peut se demander si tout événement fortuit, dans tout champ d'épreuves,

---

[1] *Voir* les notes de M de Finetti sur cette question dans les *Rend. Ist. Lombardo* de 1928, 1929, 1930 et deux notes de l'auteur dans le même périodique en 1930. *Voir* aussi, Cantelli (11).

obéit à cette loi et par suite s'il y possède une probabilité déterminée. D'autre part, quand on raisonne en imposant aux probabilités des événements considérés, des conditions (même de celles qui paraissent, au premier abord, très naturelles, comme l'invariance dans la superposition, ou comme l'égalité des probabilités des différents nombres entiers), il n'est pas certain que ces conditions soient légitimes.

Par exemple, il paraît légitime, à certains, de considérer le cas d'une variable aléatoire qui ne peut prendre qu'une infinité de valeurs numériques entières dont les probabilités respectives sont égales. Or, comme on l'a vu (p. 16), la série de ces probabilités doit être convergente. Il faut donc que leur valeur commune soit nulle. Alors la formule (14′) qui donnerait

$$1 = 0 + 0 + \ldots,$$

n'est pas applicable à ce cas. Mais, bien entendu, rien n'empêche de considérer une variable aléatoire X ne prenant qu'une suite infinie de valeurs distinctes dont les probabilités $p_1, p_2, \ldots$ vérifient la formule (14′). Dans ce cas, comme on aura

$$1 = p_1 + p_2 + \ldots + p_n + \ldots,$$

il ne sera pas possible de supposer que les probabilites $p_1, p_2, \ldots,$ soient égales.

Et, en effet, il semble précisément difficile d'imaginer des épreuves matériellement réalisables où une même variable aléatoire puisse prendre des valeurs entieres possédant un très grand nombre de chiffres aussi fréquemment que des valeurs entières usuelles.

D'ailleurs si l'on adopte la définition empirique de la probabilité développée page 5, on peut rendre vraisemblable, ainsi qu'il suit, le principe complet des probabilités totales :

Soient $E_1, E_2, \ldots, E_s, \ldots$, une suite infinie de modalités incompatibles d'un évènement fortuit E Au cours de n épreuves, ces événements se seront produits r_1 fois, r_2 fois, $\ldots$, r_s fois, $\ldots$ respectivement et pour E : $r_1 + r_2 + \ldots + r_s + \ldots = r$ fois. (Puisque $r \leqq n$, la série d'entiers $r_1 + r_2 + \ldots + r_s + \ldots$ est nécessairement convergente : ses termes sont nuls à partir d'un certain rang, variable avec le groupe de n épreuves.) On aura donc pour leurs fréquences respectives $f_k = \frac{r_k}{n}$, $f = \frac{r}{n}$, l'égalité

$$f = f_1 + f_2 + \ldots + f_s + \ldots. \tag{15}$$

Lorsqu'on change le groupe des n épreuves et le nombre des épreuves, les termes de cette égalité varient, mais l'égalité subsiste. Ces termes représentent, par définition, les mesures expérimentales des grandeurs physiques représentées par les probabilités $p, p_1, p_2, \ldots, p_s, \ldots$ des événements $E, E_1, E_2, \ldots, E_s, \ldots$ On devrait donc nécessairement conclure qu'on a

$$p = p_1 + p_2 + \ldots + p_s + \ldots \tag{16}$$

Ainsi le principe complet des probabilités totales serait nécessairement vérifié dans la classe des événements fortuits auxquels la définition empirique attribue des probabilités déterminées.

On peut évidemment élever des objections contre la démonstration précédente, en ce qui concerne la légitimité du passage de (15) à (16). Par exemple, supposons que $p = 1$ et qu'on fasse des expériences en prenant toujours pour n une puissance de 2, $n = 2^h$. Si l'on observait à chaque fois les répétitions suivantes :

$$r_1 = 2^{h-2}, \quad r_2 = 2^{h-3}, \quad \ldots, \quad r_{h-2} = 2, \quad r_{h-1} = 1,$$
$$r_h = 1 + 2^{h-1}, \quad r_{h+1} = r_{h+2} = \ldots = 0,$$

c'est-à-dire, si les fréquences étaient

$$f_1 = \frac{1}{4}, \quad f_2 = \frac{1}{8}, \quad \ldots, \quad f_{h-2} = \frac{1}{2^{h-1}},$$
$$f_{h-1} = \frac{1}{2^h}, \quad f_h = \frac{1}{2^h} + \frac{1}{2},$$
$$f_{h+1} = f_{h+2} = \ldots = 0,$$

la fréquence f_s étant toujours $\frac{1}{2^{s+1}}$ pour $n > 2^s$, on attribuerait naturellement à E_s la probabilité $p_s = \frac{1}{2^{s+1}}$ et l'on aurait

$$p_1 + p_2 + \ldots + p_s + \ldots = \frac{1}{2^2} + \frac{1}{2^3} + \ldots + \frac{1}{2^{s+1}} + \ldots = \frac{1}{2} < p.$$

Il en serait de même si, ce qui serait plus vraisemblable, les f_s éprouvaient de légères fluctuations autour des valeurs qu'on vient de leur attribuer.

En définitive, pour adopter le principe complet des probabilités totales (14'), nous nous contenterons d'observer : 1° qu'il est commode dans la théorie; 2° qu'il peut trouver place dans une

théorie non contradictoire; 3° qu'il n'est pas en contradiction avec l'expérience.

En ce qui concerne 3°, il suffit de noter qu'en pratique, on ne peut pour chaque épreuve calculer une infinité de répétitions. On ne peut qu'en calculer un grand nombre et l'on rencontre dans la vérification du principe, des difficultés analogues à celles qui se présenteraient dans l'étude beaucoup plus simple des mesures de longueurs.

Des objections pourraient, en effet, être aussi soulevées contre la propriété d'additivité complète des longueurs des intervalles.

Considérons la longueur d'un intervalle comme une constante physique dont la mesure expérimentale est fournie par les procédés ordinaires. Si l'on divise un intervalle I donné en une suite infinie d'intervalles partiels I_n, ces procédés, appliqués à chacun des intervalles I, I_n, fourniront leurs mesures respectives m, m_n. Bien que ces procédés soient basés sur la propriété d'additivité restreinte, il n'est nullement certain qu'ils fourniront l'égalité

$$m = m_1 + m_2 + \ldots + m_n + \ldots$$

Il est même certain que la vérification en est impossible, car chacune de ces mesures est obtenue par une erreur du même ordre de grandeur, de sorte que si l, l_1, l_2, ..., l_n, ..., sont les vraies longueurs et ε, ε_1, ε_2, ..., ε_n les erreurs correspondantes, il faut vérifier que

$$m + \varepsilon = m_1 + \varepsilon_1 + \ldots + m_n + \varepsilon_n + \ldots$$

ou

$$m - (m_1 + \ldots + m_n + \ldots) = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n + \ldots - \varepsilon$$

S'il n'y avait qu'un nombre fini, n, d'intervalles, il suffirait de vérifier que

$$|m - (m_1 + m_2 + \ldots + m_n)| < (n + 1)\omega,$$

ω étant l'erreur maximum à redouter sur une mesure. Lorsque la suite des intervalles est infinie, on ne voit pas bien en quoi peut consister la vérification. Et d'ailleurs, antérieurement à toute mesure, comment pourrait-on tracer mécaniquement les limites des I_n quand ceux-ci sont de rangs assez élevés ?

Conclusion. — A partir de maintenant, nous allons supposer que la probabilité est définie ainsi qu'il suit (définition axiomatique), ou qu'elle est définie de façon à vérifier (définition empirique) les conditions suivantes :

On associe à la catégorie C d'épreuves envisagées, une certaine famille $\mathcal{F}$ d'événements fortuits dits « probabilisables » (1) sur C

I. Cette famille constituera un « corps » d'événements. Autrement dit : 1° elle devra comprendre les contraires respectifs des événements qui la constituent; 2° si des événements fortuits E_1, E_2, ..., formant une suite finie ou infinie appartiennent à la famille $\mathcal{F}$, celle-ci devra contenir aussi l'événement consistant dans la réalisation de l'un au moins des événements E_1, E_2, ... (Il résulte de ces deux conditions que $\mathcal{F}$ comprend la certitude et l'impossibilité.)

II. A chaque événement E de $\mathcal{F}$, un nombre $p(E) \geq 0$, appelé probabilité de E sur C est attaché de sorte que, avec les notations du § I précédent, l'on ait, si E_1, E_2, ..., sont incompatibles,

$$(14') \qquad p(E_1) + p(E_2) + \ldots + p(E_n) + \ldots = p(E_1 \text{ ou } E_2 \text{ ou } \ldots \text{ ou } E_n \ldots)$$

et que la probabilité de la certitude soit égale à l'unité.

On déduit de I et II les conséquences suivantes. D'abord, en ce qui concerne la famille $\mathcal{F}$, si l'on considère une suite finie ou infinie d'événements H_1, H_2, ..., H_n, ... de $\mathcal{F}$, l'événement K contraire au concours H de H_1, H_2, ..., H_n, ..., consiste dans la réalisation du contraire de H_1 ou du contraire de H_2, ..., ou du contraire de H_n, ... Par suite K et alors aussi H appartiennent à $\mathcal{F}$.

Alors le concours de H_1 et du contraire de H_2, c'est-à-dire l'événement consistant en ce que H_1 ait lieu sans qu'ait lieu H_2, appartiendra aussi à $\mathcal{F}$. Plus généralement appartiendra aussi à $\mathcal{F}$ l'événement H'_1 consistant en ce que parmi H_1, H_2, ..., H_n, ..., seul H_1 ait lieu, puisqu'il consiste dans la réalisation du concours de H_1 et des contraires respectifs de H_2, de H_3, ..., de H_n, ... Il en sera de même de l'événement consistant en ce que parmi les événements H_1, H_2, ..., H_n, ... un seul d'entre eux ait lieu, car c'est l'événement (H'_1, ou H'_2, ou H'_3, ..., ou H'_n, ...).

Probabilités limites. — Passons aux relations entre les probabilités de ces divers événements.

---

(1) Par analogie avec l'expression « mesurable »

Nous venons de voir que si des événements fortuits H_1, H_2, ..., H_n, ... en suite finie ou infinie, mais dénombrable, appartiennent à $\mathcal{F}$, il en est de même du concours de ces événements

Supposons d'abord ces événements deux à deux indépendants. Alors avec les notations précédentes, K est aussi l'événement : K_1 ou K'_2 ou .. ou K'_n ou ... (en désignant par K_n le contraire de H_n et par K'_n le concours des événements indépendants H_1, H_2, . ., H_{n-1}, K_n). Si donc p est la probabilité de H, on a

$$1-p = \Pr.\{K_1\} + \Pr.\{K'_2\} + . \quad + \Pr\{K'_n\} + \quad .$$
$$= \lim_{n\to\infty} [(1-p_1) + p_1(1-p_2) + \quad + p_1 p_2 \quad p_{n-1}(1-p_n)]$$

D'où

$$p = \lim_{n\to\infty} p_1 p_2 \quad p_n \tag{17}$$

Ainsi, *lorsque des événements fortuits* H_1, H_2, . , H_n ...., appartenant à $\mathcal{F}$ *sont indépendants*, non seulement *le produit des probabilités* d'un nombre fini d'entre eux, H_1, . . ., H_n est égal à la probabilité du concours de ceux-ci, mais encore, lorsque n croît indéfiniment, *ce produit converge vers une limite qui est égale à la probabilité du concours* de cette infinité d'événements.

Considérons maintenant le cas où H_1, H_2, . . , H_n, . ., ne sont pas supposés indépendants. Alors on ramène au cas précédent en observant que la probabilité de K_n égale à $p_1 p'_2 \ldots p'_{n-1} (1-p'_n)$ en désignant en général par p'_n la probabilité de H_n dans la catégorie C'_n constituée par celle des épreuves de C où ont lieu H_1, H_2, . ., H_{n-1}. On aura alors, par le même raisonnement que plus haut,

$$p = \lim_{n\to\infty} p_1 p'_2 \cdot \quad p'_n \tag{18}$$

Dans les deux cas, on aura donc

$$\Pr\{H_1 \text{ et } H_2 . \quad \text{et } H_n . \quad\} = \lim_{n\to\infty} \Pr\{H_1 \text{ et } H_2 \quad \text{et } H_n\} \tag{19}$$

Passons maintenant à l'extension des inégalités telles que (10) et (13).

Considérons des événements H_1, H_2, . ., H_n, ..., en nombre infini, compatibles ou non, dépendants ou non, et remarquons que les deux événements

$$H_1 \text{ ou } H_2 \text{ ou } .. \text{ ou } H_n . \;.$$

et

$$H_1 \text{ ou } \{C(H_1) \text{ et } H_2\} \text{ ou } \{C(H_1) \text{ et } C(H_2) \text{ ou } H_3\} \text{ ou }$$
$$\{C(H_1) \text{ et } C(H_2) \ldots \text{ et } C(H_{n-1}) \text{ et } H_n\} \text{ ou}$$

sont identiques. Comme le second est défini à partir d'événements incompatibles, la probabilité P du premier est la somme de la série des probabilités de ces événements incompatibles. Or la somme des n premiers n'est autre que la probabilité de l'événement H_1 ou $H_2 \ldots$ ou H_n. On a donc, dans tous les cas où $H_1, H_2, \ldots,$ appartiennent à la famille $\mathcal{F}$

$$(20) \quad \Pr\{H_1 \text{ ou } H_2 \ldots \text{ ou } H_n \text{ ou } \ldots\} = \lim_{n\to\infty} \Pr.\{H_1 \text{ ou } H_2 \ldots \text{ ou } H_n\}$$

Il est parfois commode d'employer, pour les événements, une notation empruntée à la théorie des ensembles. Si l'événement E implique l'événement F, l'ensemble des épreuves où E a lieu est compris dans l'ensemble des épreuves où F a lieu, et l'on peut écrire

$$E \subseteq F$$

Soit une suite d'événements $E_1, E_2, \ldots,$ formant une suite ascendante, c'est-à-dire telle que $E_1 \subset E_2 \subset \ldots$, et soit E l'événement consistant dans la réalisation de l'un au moins des événements E_i. Alors, on a la formule utile

$$(20\ bis) \qquad \Pr E = \lim_{n\to\infty} \Pr.E_n$$

qui résulte de (20) puisqu'ici (E_1 ou $E_2, \ldots,$ ou E_n) c'est E_n.

On établit de la même façon en partant de (19) que si $F_1, F_2, \ldots,$ forme une suite descendante d'événements, c'est-à-dire telle que

$$F_1 \supset F_2 \supset F_3 \supset$$

et si F désigne le concours de ces événements, on a

$$(19\ bis) \qquad \Pr F = \lim_{n\to\infty} \Pr F_n.$$

Quand F est impossible, l'égalité (19 *bis*) constitue « l'axiome de continuité » de M. Kolmogoroff (4, p. 13) équivalent à (14′).

Quelques inégalités. — Au moyen de (20), on déduit immédiatement de (10) son extension à une suite infinie d'événements :

$$(21) \quad \Pr.\{H_i\} \leqq \Pr.\{H_1 \text{ ou } H_2 \ldots \text{ ou } H_n \text{ ou } \ldots\}$$
$$\leqq \Pr\{H_1\} + \Pr\{H_2\} + \ldots + \Pr.\{H_n\} + \ldots \qquad (i = 1, 2, \ldots, n, \ldots)$$

ou plus précisément

$$(22) \qquad \varpi(p_1, p_2, \ldots, p_n, \ldots) \leqq \mathrm{P} \leqq \Pi(p_1, p_2, \ldots, p_n, \ldots),$$

où $\varpi(p_1, p_2, \ldots, p_n, \ldots)$ est la borne supérieure de l'ensemble des p_i et où $\Pi(p_1, p_2, \ldots, p_n, \ldots)$ est le plus petit des deux nombres dont le premier est 1 et dont l'autre est la somme finie ou infinie de la série à termes non négatifs

$$p_1 + p_2 + \ldots + p_n + \ldots$$

On obtiendra de même l'extension de (13) au moyen du passage à la limite (19 *bis*). On aura *la formule de Boole généralisée* :

$$(23) \quad 1 - [\mathrm{Pr}\{\mathrm{C}(\mathrm{H}_1)\} + \ldots + \mathrm{Pr}.\{\mathrm{C}(\mathrm{H}_n)\} + \ldots] \leqq \mathrm{Pr}\{\mathrm{H}_1 \text{ et } \ldots \mathrm{H}_n \text{ et } \ldots\} \leqq \mathrm{Pr}\{\mathrm{H}_i\} \quad (i = 1, 2, \ldots, n, \ldots)$$

ou plus précisément

$$(24) \qquad \theta(p_1, p_2, \ldots, p_n, \ldots) \leqq p \leqq \Theta(p_1, p_2, \ldots, p_n, \ldots),$$

en appelant $\theta(p_1, p_2, \ldots, p_n, \ldots)$ le plus grand des deux nombres 0 et $1 - (q_1 + q_2 + \ldots)$, avec $q_i = 1 - p_i$ et en désignant par

$$\Theta(p_1, p_2, \ldots, p_n, \ldots)$$

la borne inférieure de l'ensemble des nombres $p_1, p_2, \ldots, p_n,$

Observons aussi que *les systèmes d'inégalités* (22) *et* (24) *ne peuvent être précisés* quand on y prend pour $p_1, p_2, \ldots$ des probabilités arbitraires. Autrement dit, il est impossible qu'il existe une suite infinie d'événements $\mathrm{H}_1, \mathrm{H}_2, \ldots$ de probabilités $p_1, p_2, \ldots$ telle que

$$(25) \qquad \mathrm{P} < \varpi(p_1, p_2, \ldots, p_n, \ldots)$$

ou

$$(26) \qquad \Pi(p_1, p_2, \ldots, p_n, \ldots) < \mathrm{P}.$$

Car il existerait alors un entier N tel que l'on ait, dans le cas de (25),

$$\mathrm{P} < \varpi(p_1, p_2, \ldots, p_N)$$

et, dans le cas de (26), en vertu de (20),

$$\Pi(p_1, p_2, \ldots, p_n, \ldots) < \mathrm{Pr.}(\mathrm{H}_1 \text{ ou } \mathrm{H}_2 \ldots \text{ ou } \mathrm{H}_n) \qquad \text{pour } n > \mathrm{N}.$$

On aurait alors, pour $n > N$: dans le premier cas

$$\text{Pr.}\{H_1 \text{ ou } H_2 \ldots \text{ ou } H_n\} < \varpi(p_1, \ldots, p_N) \leqq \varpi(p_1, p_2, \ldots, p_n)$$

et dans le second

$$\Pi(p_1, p_2, \ldots, p_n) \leqq \Pi(p_1, p_2, \ldots, p_n, \ldots) < \text{Pr.}\{H_1 \text{ ou } H_2 \ldots \text{ ou } H_n\},$$

inégalités également impossibles comme nous l'avons vu page 14.

On peut observer que si p_n tend vers zéro avec $\frac{1}{n}$ (et, en particulier, si la série Σp_i converge), la probabilité du concours de tous les H_i est nulle d'après (23). Si p_n tend vers 1 (et si, en particulier, la série Σq_i converge), la probabilité de la réalisation de l'un au moins des H_i est égale à l'unité d'après (21). D'ailleurs, ces deux séries ne peuvent être convergentes en même temps.

Théorèmes de Borel et de Cantelli. — M. Cantelli (5, p. 10) a déduit de la formule de Boole généralisée un résultat plus précis que les précédents en l'appliquant à la suite partielle $H_n, H_{n+1}, \ldots$. On a alors

$$1 - \text{Pr.}\{H_n \text{ et } H_{n+1} \text{ et } \ldots\} \leqq \sum_{i \geqq n} q_i$$

ou

$$\text{Pr.}\{C(H_n) \text{ ou } C(H_{n+1}) \text{ ou } \ldots\} \leqq \sum_{i \geqq n} q_i$$

Le premier membre est évidemment au moins égal à la probabilité ρ pour qu'une infinité des événements $C(H_1), C(H_2), \ldots, C(H_n), \ldots$ se produisent. On a donc

$$0 \leqq \rho \leqq \sum_{i \geqq n} q_i,$$

où ρ est indépendant de n. D'où $\rho = 0$, si Σq_i converge. Dès lors, en posant $E_n = C(H_n)$:

Si la série des probabilités d'une suite d'événements E_n (définis sur la même catégorie d'épreuves) est convergente, il y a une probabilité nulle que se produisent une infinité de ces événements dans la même épreuve. Ou encore :

Si la série des probabilités d'un même événement, au cours d'une suite infinie d'épreuves, est convergente, il y a une probabilité

nulle que cet événement se reproduise une infinité de fois au cours de cette suite d'épreuves.

On voit que M. Cantelli a pu ainsi étendre au cas d'événements dépendants une importante proposition démontrée d'abord par M. Borel (1, p. 6, *voir* aussi II, 1, p. 14) dans le cas d'événements indépendants.

M. Cantelli (5) a aussi donné une démonstration plus simple d'un second théorème démontré par M. Borel en même temps que le premier. Si, en effet, les H_n sont indépendants, alors d'apres (17) appliqué encore à la suite partielle H_n, H_{n+1}, . . , on a

$$\text{Pr}\ \{ H_n \text{ et } H_{n+1} \text{ et } \quad \} = p_n p_{n+1} \quad = (1 - q_n)(1 - q_{n+1})$$

Quand Σq_n diverge, le produit infini du second membre converge vers zéro.

Donc l'événement G_n consistant en ce que H_n et H_{n+1} et ont tous lieu à la fois a une probabilité nulle. En vertu de (21), il en sera donc de même de l'événement G_1 ou G_2 ou G_3, etc. Un tel événement consiste en ce que pour au moins une valeur de m, G_m a lieu c'est-a-dire qu'aucun des événements

$$E_m = C(H_m), \qquad E_{m+1} = C(H_{m+1})$$

n'a lieu. C'est-a-dire encore qu'il y a 0 ou seulement un nombre fini d'événements E_n se réalisant Alors il y a une probabilité égale à 1 que se réalisent une infinité d'événements E_n. D'où le théorème de Borel :

Si des événements E_1, E_2, . . . *sont* INDÉPENDANTS, *la probabilité que se réalise une infinité de ces événements ne peut être égale qu'à zéro ou à l'unité. Ces deux alternatives correspondent respectivement, la première à la convergence, la seconde à la divergence de la série*

$$(\text{Pr } E_1) + (\text{Pr. } E_2) + \quad + (\text{Pr } E_n) + . \quad . \tag{27}$$

Événements négligeables. — Nous aurons fréquemment à considérer dans le Chapitre V des événements $\mathcal{C}$ et $\mathcal{J}$ ayant respectivement la probabilité de la certitude : soit l'unité, et la probabilité de l'impossibilité : soit zéro, sans être eux-mêmes ni certains ni impossibles.

Il est important de noter que *la probabilité d'un évènement* E *n'est pas modifiée quand on néglige un évènement de probabilité nulle.* D'une manière précise, soit E_1 l'événement (E ou $\mathcal{J}$), E_2 l'événement (E sans $\mathcal{J}$). On a

$$\text{Pr. } E_1 = \text{Pr. } E_2.$$

En effet, E_1 est réalisé par E_2 ou par l'événement $e =$ (E avec $\mathcal{J}$). On a donc

$$\text{Pr } E_1 = \text{Pr } E_2 + \text{Pr } e,$$

or, $e \subset \mathcal{J}$, donc

$$\text{Pr } e \leqq \text{Pr. } \mathcal{J} = 0,$$

d'où

$$\text{Pr } e = 0 \quad \text{et} \quad \text{Pr } E_1 = \text{Pr } E_2$$

De même, E_1 est réalisé par E ou par $e' =$ ($\mathcal{J}$ sans E). Donc

$$\text{Pr. } E_1 = \text{Pr } E + \text{Pr } e'.$$

Or

$$\text{Pr. } e' \leqq \text{Pr. } \mathcal{J} = 0,$$

d'où

$$\text{Pr. } e' = 0 \quad \text{et} \quad \text{Pr } E_1 = \text{Pr. } E.$$

Finalement

$$\text{Pr. (E ou } \mathcal{J}) = \text{Pr. E} = \text{Pr (E sans } \mathcal{J}) \tag{28}$$

Par contre

$$\text{Pr. (E avec } \mathcal{J}) = 0 \tag{29}$$

Observons que (E avec $\mathcal{C}$) n'est autre qu'un (E sans $\mathcal{J}$). Donc

$$\text{Pr. (E avec } \mathcal{C}) = \text{Pr E} \tag{30}$$

SECONDE PARTIE.

LES VARIABLES ALÉATOIRES.

CHAPITRE III.

VALEURS MOYENNES DES VARIABLES ALÉATOIRES

SECTION I INTRODUCTION

Définition des variables aléatoires — Considérons une certaine catégorie C d'épreuves et supposons que le résultat R de chaque épreuve soit déterminé par le hasard. Nous appellerons *variable aléatoire* définie sur C, une variable X dont la valeur numérique soit parfaitement déterminée par le résultat R. Par exemple, C est la catégorie des jets de dé, le résultat R de chaque épreuve (consistant à jeter un dé) est la position du dé après le jet, et la valeur de X est le nombre de points visibles sur la face supérieure du dé, quand R est connu. Dans la même catégorie d'épreuves d'autres variables aléatoires pourraient être définies. Par exemple, en rapportant la position du centre du dé à deux axes horizontaux, X pourrait être l'abscisse finale de ce centre.

Remarque. — On pourrait concevoir des variables aléatoires plus générales qui ne seraient déterminées que par les résultats de *certaines* des épreuves. Par exemple, si chaque épreuve consistait à prendre au hasard un nombre y entre zéro et un et si l'expression décimale de y était la suite de chiffres

$$0, \; y_1 y_2 y_3 \ldots y_n \ldots,$$

on pourrait prendre

$$X = y_1 - y_2 + y_3 - y_4 + \dots + y_{2p-1} - y_{2p} + \dots$$

X serait déterminé et fini pour certaines épreuves, par exemple par celles qui donneraient pour y une fonction décimale exacte;

X pourrait être infini, par exemple quand

$$y = 0,1010 \dots 10 \dots,$$

X pourrait être indéterminé, par exemple, quand

$$y = 0,11 \dots 11 \dots$$

Il n'y aura pas d'inconvénient à considérer aussi de telles variables aléatoires X si l'on a soin de se borner aux cas où la probabilité que X ne soit pas déterminé et fini est égale à zéro.

D'après la définition d'une variable aléatoire X, celle-ci est une « fonctionnelle » X[R] dont « l'argument » est le résultat fortuit R d'une épreuve appartenant à une certaine catégorie. A ce titre, l'étude des variables aléatoires rentre dans « l'Analyse générale », théorie qui se propose, en particulier, de déterminer les propriétés des fonctionnelles dont l'argument est un élément de nature quelconque.

Mais si l'on se place au point de vue spécial des probabilités, il n'est toujours pas nécessaire de connaître X[R], il suffit souvent de connaître une fonction ordinaire convenablement liée à X. Cette fonction, qui a été considérée depuis longtemps mais dont l'importance n'a été reconnue que tout récemment, a été nommée par M. von Mises, fonction de répartition et par M. Paul Lévy, fonction des probabilités totales.

Fonction des probabilités totales. Premières propriétés. — Dans la suite, nous ne nous occuperons que des variables aléatoires X telles que, pour tout nombre certain x, il y ait une probabilité déterminée — désignée par la notation Pr. $[X < x]$ — que X soit inférieur à x [1]. Cette probabilité est donc une fonction déterminée de x, soit $C(x)$, que nous appellerons *fonction des probabilités totales* de

---

[1] Dans ce Traité, I, 1, p. 107, I, 2, p. 11, on a posé $C(x) = \text{Pr.}\{X \leq x\}$. Nous préférons exclure le signe d'égalité en vue du cas, important dans les applications, où X est un écart $\geqq 0$: avec notre définition, on est alors sûr que $C(0) = 0$.

la variable aléatoire X. Il est clair qu'on a

$$0 \leqq C(x) \leqq 1$$

et que cette fonction, en vertu du théorème des probabilités totales, est non décroissante. A ce titre, la fonction des probabilités totales relève de la théorie des fonctions dites *monotones*. Dans ce qui suit, nous supposerons connue cette théorie. Mais, à toutes fins utiles, nous en avons rappelés ceux des résultats qui seront utilisés ici, dans le *Supplément mathématique*, à la fin de ce volume, p. 270.

Lorsque X reste fini pour chaque épreuve possible (mais non nécessairement borné pour l'ensemble des épreuves), alors on a, en vertu de (14'),

$$\begin{aligned}\Pr \{X < a\} \\ = \Pr \{a_1 \leqq X < a\} + \Pr. \{a_2 \leqq X < a_1\} + \ldots + \Pr \{a_n \leqq X < a_{n-1}\} + \ldots\end{aligned}$$

où $a > a_1 > a_2 > \ldots > a_n \ldots$, et où a_n tend vers $-\infty$. D'où

$$C(a) = \lim_{a_n \to -\infty} [C(a) - C(a_n)],$$

c'est-à-dire

$$\lim_{a_n \to -\infty} C(a_n) = 0$$

D'une façon analogue, on verrait que

$$\lim_{b_n \to +\infty} C(b_n) = 1$$

Ainsi, lorsqu'une variable aléatoire X reste toujours finie, on a

$$(31) \qquad \boxed{\lim_{x \to -\infty} C(x) = 0, \qquad \lim_{x \to +\infty} C(x) = 1,}$$

ce que nous pourrons exprimer en disant que $C(x)$ *varie effectivement* de 0 à 1, ou représenter par les notations

$$C(-\infty) = 0, \qquad C(+\infty) = 1.$$

Dans le cas contraire, il y aurait une probabilité non nulle que $|X|$ soit infini.

D'autre part, on a, pour tout nombre a,

$$\begin{aligned}\Pr. \{a \leqq X < a_0\} \\ = \Pr. \{X = a\} + \Pr. \{a_1 \leqq X < a_0\} + \Pr. \{a_2 \leqq X < a_1\} + \ldots \\ + \Pr. \{a_n \leqq X < a_{n-1}\} + \ldots,\end{aligned}$$

où

$$a = \lim_{n \to \infty} a_n < a_{n-1} < \cdots < a_1 < a_0.$$

D'où

$$C(a_0) - C(a) = \Pr.\{X = a\} + \lim_{n \to \infty}\{C(a_0) - C(a_n)\},$$

c'est-à-dire (*voir*, au besoin, le *Supplément mathématique*, p. 270)

$$(32) \qquad C(a+0) - C(a) = \Pr.\{X = a\}.$$

Par un raisonnement analogue, on verrait que

$$(33) \qquad C(a) - C(a-0) = 0.$$

Ainsi *toute fonction des probabilités totales est continue à gauche*, et *elle n'est discontinue que pour les valeurs a dont la probabilité n'est pas nulle. La probabilité d'une telle valeur est égale au « saut »* (de même, p. 270) *de* $C(x)$ *pour* $x = a$.

Une fonction non décroissante ne pouvant avoir qu'une suite finie ou dénombrable (p. 270) de discontinuités, la probabilité que $X = a$ n'est différente de zéro que pour une suite finie ou dénombrable de valeurs de a.

Il y a lieu d'observer que les formules (31), (32) et (33) ont été établies au moyen d'un raisonnement supposant essentiellement que les événements analogues à $a < X < b$ sont probabilisables (p. 22); et en particulier que leurs probabilités vérifient la condition du type (14').

On sait que si une fonction $f(x)$ est continue sur la droite illimitée elle n'est pas toujours — témoin la fonction x^2 — uniformément continue. Mais cette singularité disparaît si l'on suppose que $f(+\infty)$ et $f(-\infty)$ existent, c'est-à-dire si $f(x)$ a une limite finie déterminée quand x tend vers $+\infty$ et une limite finie déterminée quand x tend vers $-\infty$. On le voit en observant qu'en dehors d'un intervalle assez grand dans les deux directions, l'oscillation de $f(x)$ sera très petite et que dans cet intervalle $f(x)$ sera uniformément continue.

Or la fonction de probabilité totale d'une valeur aléatoire toujours finie tend vers zéro quand x tend vers $-\infty$ et vers un quand x tend vers $+\infty$. Donc :

Si une fonction des probabilités totales est partout continue, elle est uniformément continue.

Probabilité élémentaire et densité de probabilité. — Si une fonction des probabilités totales $C(x)$ est l'intégrale définie d'une fonction $\delta(x)$,

$$C(x) = \int_{-\infty}^{x} \delta(t)\,dt,$$

elle est nécessairement continue. Mais la réciproque n'est pas vraie L'intégrale $\int_{-\infty}^{x} \delta(t)\,dt$ [du moins quand on la définit au sens de M. Lebesgue (1)] n'est, en effet, pas seulement continue, elle est *absolument continue* comme l'a montré feu Vitali. C'est-à-dire que si l'on considère une suite, finie ou non, d'intervalles quelconques x_k, x'_k ne chevauchant pas, et si l'on forme la somme

$$\sum_k \int_{x_k}^{x'_k} \delta(t)\,dt$$

cette somme tend vers zéro quand la longueur totale l des intervalles considérés tend vers zéro. Et réciproquement Vitali a démontré [*voir*, par exemple, Lebesgue (I) p 188] que si

$$\sum_k |C(x'_k) - C(x_k)|$$

tend vers zéro avec cette longueur l, $C(x)$ est nécessairement l'intégrale d'une certaine fonction $\delta(x)$

On voit qu'on peut considérer $\delta(x)$ comme la densité d'une masse linéaire répandue de telle sorte que $C(x)$ soit la masse totale de la région $X < x$. On peut appeler $\delta(x)$ *la densité de probabilité* et $\delta(x)dx$ la probabilité élémentaire On démontre (Lebesgue, I, p 185) que $\delta(x)$ est la dérivée de $C(x)$ en tout point x sauf peut-être sur un ensemble de points de mesure nulle, c'est-à-dire sauf peut-être sur les points d'un ensemble qui peut être enfermé dans une suite dénombrable d'intervalles dont la longueur totale est aussi petite que l'on veut

Autres propriétés des fonctions des probabilités totales. — On trouvera, à la fin de ce livre, p. 270, un « supplément mathématique » rappelant les principales définitions et propriétés relatives aux fonctions dites « monotones ». Dans ce qui suit nous supposerons connues ces définitions et propriétés.

Appliquons celles-ci dans le cas où les fonctions monotones considérées sont des fonctions des probabilités totales Les énoncés se simplifient légèrement.

1° *Si une fonction des probabilités totales est continue, elle est uniformément continue* (p 273)

2° *D'une suite de fonctions des probabilités totales, on peut toujours extraire une suite qui converge partout* (p. 272).

3° *Si une suite de fonctions des probabilités totales* $\gamma_n(x)$ *converge partout vers une fonction continue* $\gamma(x)$ *et si cette fonction limite varie effectivement de zéro à un, la convergence de la suite est uniforme.* C'est la proposition de M. Polya établie I, I, p. 157. Cette proposition a été étendue par M. Cantelli (10, 9), puis légèrement complétée par nous (p. 276) sous la forme suivante :

3° *bis. La condition nécessaire et suffisante pour qu'une suite de fonctions des probabilités totales* $\gamma_n(x)$ *converge uniformément est : A, que* $\gamma_n(x)$ *converge quel que soit* x *vers une fonction limite* $\gamma(x)$; *B, que* $\gamma_n(x-0)$ *et* $\gamma_n(x+0)$ *convergent respectivement vers* $\gamma(x-0)$ *et* $\gamma(x+0)$; *C, que la fonction* $\gamma(x)$, *continue ou non, varie effectivement de* 0 *à* 1.

Les conditions B et C peuvent d'ailleurs aussi s'exprimer sous la forme suivante. B', *que* $\gamma_n(x+0)$ *converge vers* $\gamma(x+0)$, C', *que* $\gamma(x)$ *soit elle-même une fonction des probabilités totales*, c'est-à-dire que $\gamma(x)$ ait pour bornes 0 et 1 et soit continue à gauche.

Il y a d'ailleurs lieu d'observer qu'*une fonction*, $\gamma(x)$, *limite de fonctions des probabilités totales*, $\gamma_n(x)$, *n'est pas nécessairement elle-même fonction des probabilités totales :*

I. Cette fonction-limite peut avoir des bornes intérieures à (0,1) comme on le voit en prenant $\gamma_n(x)=\frac{1}{2}$ pour $-n \leqq x \leqq +n$. Dans ce cas $\gamma(x)$ est constamment égal à $\frac{1}{2}$.

II. La fonction-limite peut aussi être discontinue à gauche, comme on le voit en prenant pour $\gamma_n(x)$ la fonction de probabilité totale d'une variable aléatoire X_n qui ne peut prendre que la valeur $-\frac{1}{n}$. Alors $\gamma(x)=0$ pour $x<0$, $=1$ pour $x \geqq 0$.

SECTION II VALEURS MOYENNES

Définitions. — On a déjà défini (I, 2, p 13) la valeur moyenne $\bar{X}$ (ou $\mathfrak{M}X$) d'un nombre aléatoire X comme la valeur de l'intégrale

$$(34) \qquad \boxed{\bar{X} = \int_{-\infty}^{+\infty} x \, d\,C(x),}$$

où $C(x)$ est la fonction des probabilités totales de X, c'est-à-dire que $C(x) = \text{Pr.}\{X < x\}$. On définit généralement le second membre comme la limite de l'intégrale de Stieltjes ([1]).

$$(35) \qquad J = \int_a^b x \, d\,C(x),$$

quand a et b tendent respectivement et indépendamment vers $-\infty$ et $+\infty$. Cette dernière intégrale existe certainement, mais elle n'a pas nécessairement une limite quand a et b tendent vers $-\infty$ et $+\infty$.

Tel serait, par exemple, le cas ou

$$C(x) = \frac{1}{2} + \frac{1}{\pi} \text{arc tang}\, x$$

et où par suite

$$J = \int_a^b x \, d\,C(x) = \frac{1}{2\pi} \, l \frac{1+b^2}{1+a^2}$$

Alors on peut faire tendre J vers n'importe quelle valeur positive ou négative ou vers l'infini en faisant tendre simultanément de manière convenable, a et b vers $-\infty$ et $+\infty$.

Ainsi la valeur moyenne d'un nombre aléatoire X n'est nécessairement ni finie ni déterminée. Mais elle l'est évidemment dans le cas particulièrement important où X est borné, c'est-à-dire où toutes les valeurs que peut prendre X restent comprises entre deux nombres fixes.

---

(1) *Voir*, par exemple, I, I, p 111, pour la définition de cette intégrale de Stieltjes. L'emploi de l'intégrale de Stieltjes en Calcul des Probabilités semble dû simultanément à M. von Mises (1, p. 20 et 72) et à M Hagstroem (1).

Dans le cas général où X peut n'être pas borné, on a

$$\int_a^b x\,d\mathrm{C}(x) = \int_0^b x\,d\mathrm{C}(x) + \int_a^0 x\,d\mathrm{C}(x)$$

Ces deux intégrales sont évidemment des fonctions monotones de b et de a. Pour que $\overline{\mathrm{X}}$ existe, il faut et il suffit que ces deux intégrales aient des limites finies, c'est-à-dire que les intégrales

$$\int_0^{+\infty} x\,d\mathrm{C}(x) \quad \text{et} \quad \int_{-\infty}^0 x\,d\mathrm{C}(x)$$

soient convergentes.

Alors l'intégrale $\int_{-\infty}^{+\infty} |x|\,d\mathrm{C}(x)$ sera aussi convergente, réciproquement, si cette dernière est convergente, $\int_a^b x\,d\mathrm{C}(x)$ aura une limite finie et déterminée, indépendante de la façon dont a et b tendent vers $-\infty$ et $+\infty$. Nous pourrons donc dire que pour que cette condition soit remplie, il faut et il suffit que l'intégrale $\int_{-\infty}^{+\infty} x\,d\mathrm{C}(x)$ soit absolument convergente.

Observons que, dans certains cas, il pourrait être naturel d'attribuer à X une valeur moyenne sans que cette intégrale soit absolument convergente. Ainsi, dans le cas, où les valeurs de X égales et de signe contraire sont également probables {ou plus précisément, si

$$\Pr\{a < \mathrm{X} < b\} = \Pr\{-b < \mathrm{X} < -a\},$$

quels que soient les nombres positifs a et b}, il est indiqué de considérer que X a une valeur moyenne nulle (¹). Plus généralement,

(¹) Supposons $\{\Pr[\mathrm{X} < x]\} = \int_{-\infty}^{x} \varphi(t)\,dt$ avec $\varphi(t) = \varphi(-t)$, on aura $\int_{-b}^{+b} x\,\varphi(x)\,dx = 0$, et il sera naturel de prendre $\overline{\mathrm{X}} = 0$. Si de plus $\varphi(t) = \mathrm{A}$ pour $0 \leqq t \leqq 1$ et $\varphi(t) = \mathrm{A}\,t^{-\alpha}$ pour $t \geqq 1$, l'intégrale $\int_{-\infty}^{+\infty} \varphi(x)\,dx$ sera absolument convergente pour $\alpha > 1$ et égale à 1, pour $\mathrm{A} = \frac{\alpha - 1}{2\alpha}$. Pourtant, si $1 < \alpha < 2$, l'intégrale $\int_{-\infty}^{+\infty} |x|\,\varphi(x)\,dx$ sera divergente, et, d'après la première définition, $\overline{\mathrm{X}}$ n'aurait pas de sens.

on pourrait poser

$$X = \lim_{b \to +\infty} \int_{-b}^{+b} x \, d\,C(x)$$

Mais, il y aurait à cela quelques inconvénients (*voir* p. 45), de sorte que nous nous en tiendrons à notre première définition.

Si, au sens de cette première définition, $\overline{X}$ existe, les deux intégrales

$$\int_0^{+\infty} x \, d\,C(x) \quad \text{et} \quad \int_{-\infty}^{0} x \, d\,C(x)$$

sont convergentes. Il en sera donc de même de

$$\int_0^{+\infty} x \, d[C(x) - C(-x)].$$

Or, pour $x \geqq 0$,

$$C(x) - C(-x) = \Pr\{-x \leqq X < x\}.$$

Le second membre ne diffère de la fonction

$$q(x) = \Pr\{|X| < x\} \tag{36}$$

s'il en diffère, que pour un ensemble dénombrable de valeurs de x, nécessairement positives. D'après la définition de l'intégrale de Stieltjes, les intégrales

$$\int_0^{+\infty} x \, d[C(x) - C(-x)] \quad \text{et} \quad \int_0^{+\infty} x \, d\,q(x)$$

sont alors égales. Or la seconde est par définition la valeur moyenne de $|X|$. Inversement, si la valeur moyenne de $|X|$ est finie (et alors elle a nécessairement une valeur déterminée), les deux intégrales

$$\int_0^{+\infty} x \, d\,C(x), \quad \int_0^{+\infty} x \, d[-C(-x)]$$

qui sont $\geqq 0$, seront convergentes et l'intégrale qui donne $\overline{X}$

$$\int_{-\infty}^{+\infty} x \, d\,C(x)$$

sera absolument convergente.

En résumé, pour que $\overline{X}$ ait une valeur finie et déterminée, il faut et il suffit que la valeur moyenne de $|X|$ soit finie.

Second mode de calcul de la valeur moyenne. — Nous avons défini $\overline{X}$ comme égal à $\int_{-\infty}^{+\infty} x\, d\mathrm{C}(x)$. Et nous avons déterminé cette intégrale comme la limite supposée existante et unique de $\int_a^b x\, d\mathrm{C}(x)$ lorsque a et b convergent indépendamment vers $-\infty$ et $+\infty$. Il est souvent plus commode d'utiliser une seconde définition plus directe mais équivalente :

Pour que $\overline{X}$ soit déterminé et fini ainsi que $\overline{|X|}$, il faut et il suffit qu'il existe *une* suite de nombres x_i croissant de $-\infty$ à $+\infty$, divisant la droite en segments dont les longueurs $x_i - x_{i-1}$ ont une borne supérieure finie δ, et *une* suite de nombres ξ_i sur ces segments respectifs, tels que la série

$$(37) \qquad \mathrm{S} = \sum_{i=-\infty}^{i=+\infty} \xi_i [\mathrm{C}(x_i) - \mathrm{C}(x_{i-1})] = \sum_{i=-\infty}^{i=+\infty} \xi_i \,\mathrm{Pr}\{x_{i-1} \leqq X < x_i\}$$

soit absolument convergente. Et alors, on a

$$(38) \qquad \overline{X} = \lim_{\delta \to 0} \mathrm{S},$$

quels que soient les suites des x_i et des ξ_i satisfaisant aux conditions ci-dessus.

Cette proposition peut être considérée comme cas particulier d'une proposition plus générale suivante :

Soit $\varphi(x)$ une fonction partout continue, $v(x)$ une fonction dont la variation totale $\mathrm{V}(x)$ [*voir* définition, p. 273, note [1]] à partir d'un point arbitraire fixe c est bornée sur la droite illimitée. Pour que les intégrales

$$J = \int_a^b \varphi(x)\, dv(x) \qquad \text{et} \qquad K = \int_a^b |\varphi(x)|\, d\mathrm{V}(x)$$

convergent respectivement vers des limites déterminées quand a et b tendent, indépendamment, vers $-\infty$ et $+\infty$, il faut et il suffit que, pour au moins une suite de nombres x_i croissant de $-\infty$ à $+\infty$ et tels que l'oscillation de $\varphi(x)$ dans chaque segment (x_{i-1}, x_i) ait une borne supérieure finie ω et pour au

moins une suite de nombres ξ_i pris respectivement dans ces segments, la série

$$s = \sum_{i=-\infty}^{i=+\infty} \varphi(\xi_i)[v(x_i) - v(x_{i-1})]$$

soit absolument convergente. Et alors, on aura

$$I = \lim_{\omega \to 0} s.$$

Si, en particulier, $\varphi(x)$ est uniformément continue sur toute la droite (ce qui, par exemple, a lieu pour la fonction x et non pour x^2), on pourra se contenter de prendre les $x_i - x_{i-1} < \delta$ sans s'occuper de ω et l'on aura

$$I = \lim_{\delta \to 0} s.$$

Pour abréger ce livre, nous avons publié ailleurs (Fréchet, 183) la démonstration (qui d'ailleurs est facile) des propositions précédentes. Nous avons en même temps montré que la proposition sur $\int_{-\infty}^{+\infty} \varphi(x)\, dv(x)$ cesserait d'être exacte si au lieu de supposer que la variation totale $V(x)$ est bornée sur la droite illimitée, on la supposait seulement finie en chaque point x.

Valeur médiane. — La valeur moyenne $\bar{X}$ est une valeur typique de l'ensemble des valeurs d'une variable aléatoire X. Il y a souvent intérêt à considérer d'autres valeurs typiques et en particulier la « valeur médiane » (dite aussi valeur probable) de X. C'est une valeur M dont on peut dire en gros qu'il y a autant de chances pour que X soit plus grand ou plus petit que M. Précisons :

La probabilité pour que $X > M$ est $1 - C(M+0)$ et pour que $X < M$, $C(M-0)$. S'il existe un nombre M tel que

$$(39) \qquad 1 - C(M+0) = C(M-0),$$

c'est donc une valeur médiane de X. Or on a, dans ce cas,

$$(40) \quad [\Pr\{X < M\}] + [\Pr.\{X = M\}] + [\Pr\{X > M\}] \geqq 2[\Pr.\{X < M\}].$$

D'où

$$(41) \quad [\Pr.\{X < M\}] \leqq \frac{1}{2} \qquad \text{et, de même,} \qquad [\Pr.\{X > M\}] \leqq \frac{1}{2}.$$

Il n'y a pas toujours un nombre M vérifiant (39). Par exemple, on peut avoir

$$C(a-0)=\frac{1}{3}=C(a) \qquad C(a+0)=\frac{3}{4}.$$

Le nombre a est évidemment le seul qui, substitué à M, vérifiera les deux inégalités (41); c'est donc le seul qui pourrait vérifier (39) et il ne la vérifie pas. Il faut donc donner une définition un peu moins stricte de la médiane. Nous appellerons valeur médiane *tout nombre* M *vérifiant simultanément les deux inégalités* (41).

Si la probabilité de ce nombre est nulle, alors en vertu de (41), on aura

$$\Pr\{X<M\}=\Pr\{X>M\}=\frac{1}{2}.$$

Dans le cas contraire, l'une au moins de ces probabilités est $<\frac{1}{2}$.

Il reste à montrer qu'il existe au moins une valeur médiane. Soit M' la borne supérieure des nombres α tels que

$$\left[\Pr\{X<\alpha\}\right]<\frac{1}{2},$$

M'' la borne inférieure des nombres β tels que

$$\left[\Pr\{X>\beta\}\right]<\frac{1}{2};$$

tout α est au plus égal à tout β, donc $M'\leqq M''$. On a

$$C(M'-0)\leqq\frac{1}{2}\leqq C(M''+0).$$

Donc

$$C(M'-0)\leqq\frac{1}{2}\leqq C(M'+0)\leqq C(M''-0)\leqq\frac{1}{2}\leqq C(M''+0),$$

d'où

$$C(M'-0)\leqq\frac{1}{2}=C(M'+0)=C(M''-0)=\frac{1}{2}\leqq C(M''+0).$$

Dès lors si

$$M'\leqq M\leqq M'',$$

on a

$$(42)\qquad C(M-0)\leqq\frac{1}{2}\leqq C(M+0),$$

c'est-à-dire que les inégalités (41) sont vérifiées. Ainsi, *il y a toujours au moins une valeur médiane* M *de* X; *s'il y en a plus d'une, elles remplissent un segment* $M'\leqq M\leqq M''$, et pour toute valeur

intérieure à ce segment, on a

$$\Pr\{X < M\} = \Pr\{X > M\} = \frac{1}{2}.$$

Quartiles. — On définit souvent aussi un premier et troisième « quartiles », le second quartile se confondant avec la médiane. Ce sont des valeurs M_1, M_3 telles que

$$\Pr\{X < M_1\} \leqq \frac{1}{4}, \qquad \Pr\{X > M_1\} \geqq \frac{3}{4},$$

$$\Pr\{X < M_3\} \leqq \frac{3}{4}, \qquad \Pr\{X > M_3\} \geqq \frac{1}{4}.$$

On démontre, comme pour la médiane, l'existence d'au moins un premier, et d'au moins un troisième quartile.

Valeur moyenne d'une somme. — Soient deux variables aléatoires X et Y. Si l'on connaît leurs valeurs moyennes $\overline{X}$, $\overline{Y}$, peut-on déterminer celle de leur somme $Z = X + Y$? Pour que le problème ait un sens, il faut naturellement supposer d'abord que X et Y soient définis sur la même catégorie d'épreuves, ou alors ne définir Z que sur les épreuves où X et Y sont à la fois définies. Dans ce dernier cas, on devra aussi supposer que les moyennes de X et de Y ne sont calculées que sur cet ensemble d'épreuves communes.

On peut aussi supposer que, X et Y étant respectivement définies sur les catégories initiales C_1 et C_2, on définisse Z sur la catégorie complexe (*voir* p. 9) issue de C_1 et C_2.

I. Si les deux variables aléatoires X, Y ne peuvent prendre qu'un nombre fini de valeurs distinctes

$$x_1, \quad x_2, \quad \ldots, \quad x_n; \qquad y_1, \quad y_2, \quad \ldots, \quad y_r$$

avec les probabilités respectives

$$p'_1, \quad p'_2, \quad \ldots, \quad p'_n, \qquad p''_1, \quad p''_2, \quad \ldots, \quad p''_r,$$

leur somme $Z = X + Y$ ne peut prendre aussi qu'un nombre fini de valeurs distinctes $z_1, \ldots, z_s$ de la forme $z_k = x_i + y_j$. La probabilité que $Z = z_k$ est la probabilité $p_{i,j}$ pour que $Z = x_i$ et $Y = y_j$ simultanément ou, plus précisément, la somme

$$p_k = \sum_{x_i + y_j = z_k} p_{i,j}$$

des probabilités $p_{i,j}$ correspondant aux couples x_i, y_j dont la somme est égale à z_k, s'il y en a plusieurs. On a alors par définition,

$$\overline{Z} = \sum_k p_k z_k = \sum_k \left\{ \sum_{x_i + y_j = z_k} p_{i,j}(x_i + y_j) \right\}$$
$$= \sum_{i,j} p_{i,j}(x_i + y_j)$$
$$= \sum_i x_i \left[\sum_j p_{i,j} \right] + \sum_j y_j \left[\sum_i p_{i,j} \right]$$
$$= \sum_i x_i p'_i + \sum_j y_j p''_j$$

D'où

$$\overline{X + Y} = \overline{X} + \overline{Y}$$

II Cette formule subsiste dans le cas général, mais la démonstration en est un peu moins simple, d'autant qu'il s'y introduit la question de l'existence des intégrales qui, dans ce cas, représentent les moyennes considérées.

Supposons donc que $\overline{X}$ et $\overline{Y}$ existent, c'est-à-dire que les intégrales qui les définissent

$$\overline{X} = \int_{-\infty}^{+\infty} x \, d\{\text{Pr.}[X \leq x]\},$$
$$\overline{Y} = \int_{-\infty}^{+\infty} y \, d\{\text{Pr.}[Y \leq y]\}$$

soient absolument convergentes. Soient alors

$$\ldots, x_{-1}, x_0, x_1, \ldots \qquad \ldots, y_{-1}, y_0, y_1$$

deux suites croissant de $-\infty$ et $+\infty$, avec l'équidistance constante

$$x_i - x_{i-1} = y_j - y_{j-1} = \delta$$

Alors, d'apres la p. 38,

$$(42\ bis) \quad \begin{cases} \overline{X} = \sum_{i=-\infty}^{i=+\infty} \xi_i \{\text{Pr.}[x_{i-1} \leqq X \leq x_i]\}, \\ \overline{Y} = \sum_{j=-\infty}^{j=+\infty} \eta_j \{\text{Pr.}[y_{j-1} \leqq Y \leq y_j]\}, \end{cases}$$

en désignant par ξ_i, η_j des nombres convenablement choisis dans les

intervalles correspondants. En vue d'écrire les deux termes sous une forme analogue, observons qu'on a

$$\text{Pr.}|x_{i-1} \leqq X < x_i| = \sum_{j=-\infty}^{j=+\infty} p_{i,j},$$

$$\text{Pr.}|y_{j-1} \leqq Y < y_j| = \sum_{i=-\infty}^{i=+\infty} p_{i,j}$$

avec

$$p_{i,j} = \text{Pr.}|x_{i-1} \leqq X < x_i \text{ et } y_{j-1} \leqq Y < y_j|.$$

Et comme les différentes séries considérées sont absolument convergentes (p. 38), on a

$$\overline{X} + \overline{Y} = \sum_{i=-\infty,\ j=-\infty}^{i=+\infty,\ j=+\infty} (\xi_i + \eta_j) p_{i,j}$$

Prenons en particulier

$$x_i = i\delta, \qquad y_j = j\delta, \qquad z_k = (i+j)\delta \qquad \text{avec } k = i+j.$$

Alors

$$\xi_i + \eta_j = z_{k-1} + \theta_{i,j}\delta \qquad \text{avec } |\theta_{i,j}| \leqq 1$$

Or, si l'on représente dans le plan les lignes $X = x_i$, $Y = y_i$,

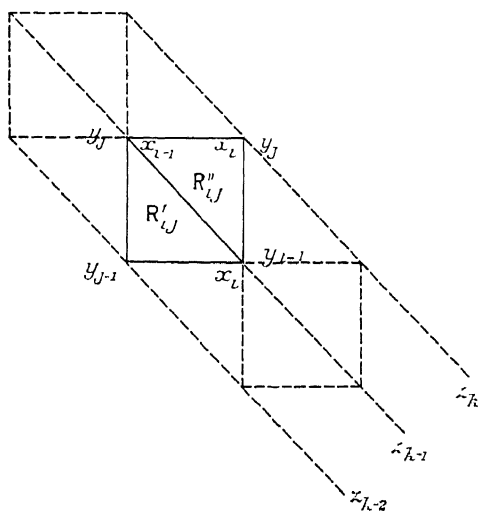



Fig. 1.

$X + Y = z_k$, on voit (*fig.* 1) que $p_{i,j}$ est la probabilité pour que le

point X, Y soit dans une certaine région analogue à un carré ouvert. C'est la somme des probabilités $p'_{i,j}$, $p''_{i,j}$ pour que ce point soit dans une des deux moitiés ouvertes $R'_{i,j}$, $R''_{i,j}$ du carré divisé par la diagonale $X + Y =$ const. On a dans ce carré

$$z_{k-2} \leqq X + Y < z_k$$

De sorte que $R'_{i,j}$ est la moitié où

$$z_{k-2} \leqq X + Y < z_{k-1}$$

et $R''_{i,j}$ celle où

$$z_{k-1} \leqq X + Y < z_k$$

On a alors

$$\begin{aligned}
\overline{X + Y} &= \sum_{i,j} (\xi_i + \eta_j) p'_{i,j} + \sum_{i,j} (\xi_i + \eta_j) p''_{i,j} \\
&= \sum_{i,j} (z_{k-1} + \theta_{i,j}\delta) p'_{i,j} + \sum_{i,j} [z_k - (1 - \theta_{i,j})\delta] p''_{i,j} \\
&= \sum_k z_k \left[\sum_{i+j=k} p'''_{i,j} \right] + \Theta\delta \\
&= \sum_k z_k p_k + \Theta\delta
\end{aligned}$$

avec

$$|\Theta| = \left| \sum_{i,j} \theta_{i,j} p'_{i,j} - \sum_{i,j} (1 - \theta_{i,j}) p''_{i,j} \right| \leqq \sum_{i,j} (p'_{i,j} + p''_{i,j}) = 1,$$

ou

$$p_k = \sum_{i+j=k+1} p'''_{i,j}$$

(et où $p'''_{i,j}$ est la probabilité pour que le point X, Y soit dans un des triangles ouverts $R'_{i,j}$, avec $i + j = k + 1$; $R''_{i,j}$, avec $i + j = k$ tels que $z_{k-1} \leqq X + Y < z_k$). La série

$$S = \sum_{k=-\infty}^{k=+\infty} z_k p_k$$

est donc absolument convergente. Or, il est clair que p_k est la probabilité pour que $Z = X + Y$ se trouve dans l'une des rangées de triangles limités par $X + Y = z_{k-1}$, $X + Y = z_k$, ou plus précisément pour que $z_{k-1} \leqq Z < z_k$. Puisque cette série S est absolument convergente, il résulte de la page 38, que $\overline{Z}$ existe

Or la somme S est

$$\overline{X}+\overline{Y}-\theta\delta \qquad \text{avec } |\theta| \leq 1$$

Lorsque δ tend vers o, la somme S de cette série tend donc vers $\overline{X}+\overline{Y}$ et, par définition de $\overline{Z}$, elle tend aussi vers $\overline{Z}$.

En résumé, la propriété déjà démontrée dans le cas des probabilités discontinues se trouve maintenant étendue au cas le plus général : *si deux variables aléatoires* X, Y, *définies sur la même catégorie d'épreuves, ont des valeurs moyennes finies et déterminées* $\overline{X}$, $\overline{Y}$, *il en est de même de leur somme* $Z = X + Y$ *et l'on a*

$$\boxed{\overline{Z} = \overline{X} + \overline{Y}} \tag{43}$$

La démonstration précédente suppose essentiellement que les intégrales qui définissent $\overline{X}$ et $\overline{Y}$ sont absolument convergentes.

Et la formule peut devenir fausse quand on appelle, comme il est naturel en certains cas, valeur moyenne de X, la limite, quand elle existe, représentée par le symbole

$$\overline{\overline{X}} = \lim_{h\to+\infty} \int_{-h}^{+h} x\, d\{\Pr[X < x]\}$$

Prenons, par exemple, pour Y, un nombre certain, $a \neq 0$ et pour X une variable aléatoire telle que l'on ait $\varphi(-x) = \varphi(x)$ quand on suppose

$$\Pr.[X < x] = \int_{-\infty}^{x} \varphi(t)\,dt$$

Alors

$$\overline{\overline{X}} = 0, \qquad \overline{\overline{Y}} = a$$

et

$$\overline{\overline{X}} + \overline{\overline{Y}} = a$$

Or, si l'on pose $Z = X + Y$, on peut choisir $\varphi(x)$ de sorte que l'on n'ait pas

$$\overline{\overline{Z}} = \overline{\overline{X}} + \overline{\overline{Y}}$$

En effet on a

$$\overline{\overline{Z}} = \lim_{h\to+\infty} J_h$$

avec

$$
\begin{aligned}
J_b &= \int_{-b}^{+b} z\, d\{\Pr[X+a < z]\} \\
&= \int_{-b-a}^{b-a} (x+a)\,\varphi(x)\,dx \\
&= \int_{-b-a}^{b+a} x\,\varphi(x)\,dx - \int_{b-a}^{b+a} x\,\varphi(x)\,dx + a\int_{-b-a}^{b-a} \varphi(x)\,dx
\end{aligned}
$$

La première intégrale du dernier membre est nulle, la troisième tend vers a quand $b \to +\infty$. Il suffit d'établir que la seconde intégrale ne tend pas alors vers zéro. Or elle est égale à $-2a\xi\varphi(\xi)$ où $b-a \leqq \xi \leqq b+a$, quand $\varphi(x)$ est une fonction continue, et l'on peut former une fonction paire $\varphi(x)$, telle que $\varphi(x) \geqq 0$, que

$$\int_{-\infty}^{+\infty} \varphi(x)\,dx = 1$$

et que $\xi\varphi(\xi)$ ne tende pas vers zero quand $b \to +\infty$. On peut, par exemple, prendre pour représentation graphique de $y = \varphi(x)$ une ligne polygonale illimitée passant par tous les points d'abscisse entière de l'axe Ox sauf les sommets ayant une abscisse de la forme $\pm(p+1)^2$, $(p = 1, 2, \ldots)$ avec, en ces points $\varphi(x) = \dfrac{A}{|x|}$. L'intégrale $\int_{-\infty}^{+\infty} \varphi(x)\,dx$, égale à une série d'aires de triangles

$$2\sum_{p=1}^{p=\infty} \frac{A}{(p+1)^2},$$

peut être rendue égale à l'unité en choisissant A convenablement. Pourtant en prenant, par exemple, $a = \dfrac{1}{2}$, on aura

$$\varphi(\xi) \geqq \frac{A}{2(p+1)^2} \qquad \text{pour } b = (p+1)^2$$

et, par suite, le produit

$$\xi\,\varphi(\xi) \geqq \left[(p+1)^2 - \frac{1}{2}\right] \frac{A}{2(p+1)^2}$$

ne pourra tendre vers zéro.

Valeur moyenne d'une différence. — D'après le résultat précédent on voit que si $U = X - Y$ et si U et Y ont des valeurs moyennes finies et déterminées, X aussi et l'on a

$$\overline{U} = \overline{X} - \overline{Y}.$$

On peut même montrer que si X et Y ont des valeurs moyennes déterminées, il en est de même de U.

Il suffit de prouver que si Y a une valeur moyenne finie et déterminée, il en est de même de $-Y$ et que $(\overline{-Y}) = -\overline{Y}$. Or, appelons $B(y)$ et $B_1(y)$ les fonctions des probabilités totales de Y et de $(-Y)$. On a

$$B_1(y) = \{\Pr[-Y < y]\} = \{\Pr[Y > -y]\} = 1 - \{\Pr[Y \leqq -y]\}$$

et si y est un point de continuité de $B(y)$, on conclut de cette égalité

$$B_1(y) = 1 - B(-y)$$

Cette égalité ne pouvant cesser d'être vraie que sur un ensemble dénombrable de points y, on aura

$$\overline{Y} = \int_{-\infty}^{+\infty} y\, dB(y) = \int_{+\infty}^{-\infty} (-y)\, dB(-y)$$
$$= \int_{+\infty}^{-\infty} (-y)[-dB_1(y)] = -\int_{-\infty}^{+\infty} y\, dB_1(y)$$

Donc $(\overline{-Y})$ existe et est égal à $-\overline{Y}$.

Valeur moyenne d'une fonction donnée d'une variable aléatoire. — Si X est une variable aléatoire et $\varphi(x)$ une fonction définie pour toute valeur de x (ou tout au moins pour toutes les valeurs de x susceptibles d'être prises par X), $Y = \varphi(X)$ est aussi une variable aléatoire. Par suite, sa valeur moyenne $\overline{Y}$ doit, d'après ce qui précède, être définie par l'intégrale

$$(44) \qquad \overline{Y} = \int_{-\infty}^{+\infty} y\, d\Phi(y),$$

où

$$(45) \qquad \Phi(y) = \Pr[\varphi(X) < y].$$

Pour que $\overline{Y}$ soit déterminé, il faut d'ailleurs que cette intégrale soit absolument convergente ou encore que $|Y|$ ait aussi une valeur moyenne déterminée [1].

---

[1] En réalité, la première condition à réaliser serait *l'existence* de la probabilité $\Phi(y)$, c'est-à-dire l'existence d'une probabilité déterminée pour que X prenne

En réalité, on définit généralement $\overline{Y} = \overline{\varphi(X)}$ comme la valeur de l'intégrale

$$I = \int_{-\infty}^{+\infty} \varphi(x)\, d\mathrm{C}(x),$$

où $\mathrm{C}(x)$ est la fonction de probabilité totale de X. Cette définition coïncide avec la précédente quand $\varphi(x) \equiv x$. Pour que la définition habituelle soit non contradictoire, il faudrait donc que les expressions de $\overline{Y}$ et de I fussent toujours égales. Et d'abord, il faudrait au moins qu'elles eussent un sens en même temps. Or il n'en est pas ainsi. Lorsque $Y = \varphi(X)$ est borné dans l'ensemble des épreuves, $\overline{Y}$ est bien déterminée, même si $\varphi(x)$ est une fonction discontinue. Il n'en est pas de même pour l'intégrale I dont la définition, qui est devenue classique quand $\varphi(x)$ est continue, varie avec les auteurs quand $\varphi(x)$ est discontinue.

Même si l'on se bornait au cas où $\varphi(x)$ est une fonction continue, on ne pourrait, quand $\varphi(x)$ n'est pas bornée, accepter sans réserve la définition habituelle de $\overline{\varphi(X)}$: on a vu page 37 que si $\int_a^b y\, d\Phi(y)$ a une limite déterminée quand a et b tendent indépendamment vers $-\infty$ et $+\infty$ respectivement, alors en posant

$$Q(y) = \Pr\,[\,|\varphi(X)| < y\,]$$

pour $y \geqq 0$, la valeur moyenne de $|\varphi(X)|$, définie sous la forme

$$\overline{|\varphi(x)|} = \int_0^{+\infty} y\, d\,Q(y),$$

est aussi finie et déterminée. Et pourtant si l'intégrale $\int_{-\infty}^{+\infty} \varphi(x)\, d\mathrm{C}(x)$ a un sens, c'est-à-dire s'il y a une limite déterminée et unique de

---

l'une des valeurs de l'ensemble E_y sur lequel $\varphi(X) < y$. Si, par exemple,

$$[\Pr.\,\{X < x\}] = \mathrm{C}(x) = x \quad \text{pour} \quad 0 \leqq x \leqq 1,$$
$$\mathrm{C}(x) = 0 \quad \text{pour} \quad x \leqq 0, \qquad \mathrm{C}(x) = 1 \quad \text{pour} \quad x \geqq 1,$$

alors $\Phi(y)$ devrait être la mesure de E_y. Si $\varphi(y)$ n'était pas mesurable, on pourrait choisir φ et y de sorte que $\Phi(y)$ n'ait pas de sens. Toutefois, on voit facilement que si $\varphi(y)$ est partout continue, $\Phi(y)$ est pour chaque variable aléatoire X, bien déterminée. Nous supposerons dans la suite, X et φ tels que $\Phi(y)$ soit une fonction bien déterminée.

$\int_a^b \varphi(x)\,d\,C(x)$ quand a et b convergent indépendamment vers $-\infty$ et $+\infty$, il n'en résulte nullement, même si l'on suppose $\varphi(x)$ continue, que l'intégrale

$$\int_{-\infty}^{+\infty} |\varphi(x)|\,d\,C(x)$$

ait un sens.

Par exemple, prenons $C(x) = \frac{e^x}{e^x+1}$ qui croît de 0 à 1 quand x croît de $-\infty$ à $+\infty$. On a, en posant $u = C(x)$, d'où $x = \mathcal{L}\,\frac{u}{1-u}$,

$$\int_{-\infty}^{+\infty} \varphi(x)\,d\,C(x) = \int_0^1 f(u)\,du$$

avec

$$f(u) = \varphi\left(\mathcal{L}\,\frac{u}{1-u}\right).$$

On pourra prendre pour $f(u)$ une fonction bornée et continue dont la représentation graphique soit formée d'arcs de courbes (par exemple les côtés d'une ligne polygonale) placés alternativement au-dessus et au-dessous de l'axe des u et limitant avec Ox des aires égales à

$$1, \quad \frac{1}{2}, \quad \frac{1}{3}, \quad \ldots, \quad \frac{1}{n}, \quad \ldots$$

La fonction

$$\varphi(x) = f\left(\frac{e^x}{e^x+1}\right)$$

sera continue pour tout x fini et les intégrales

$$\int_{-\infty}^{+\infty} \varphi(x)\,d\,C(x), \qquad \int_{-\infty}^{+\infty} |\varphi(x)|\,d\,C(x),$$

limites des sommes des n premiers termes des séries

$$1 - \frac{1}{2} + \frac{1}{3} - \ldots, \qquad 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots$$

seront la première finie, la seconde infinie.

Et pourtant, il est bien clair que, dans cet exemple, $\int_a^b \varphi(x)\,d\,C(x)$ aura une limite déterminée indépendante des modes de croissance respectifs de a et de b vers $-\infty$ et $+\infty$

Par contre, l'existence de

$$\int_{-\infty}^{+\infty} |\varphi(x)|\, d\mathrm{G}(x)$$

qui assure celle de

$$\int_{-\infty}^{+\infty} \varphi(x)\, d\mathrm{G}(x),$$

suffit pour assurer l'existence de

$$\int_{-\infty}^{+\infty} y\, d\Phi(y)$$

et réciproquement, comme nous allons le montrer.

I. Nous considérerons d'abord le cas particulier où $\varphi(x)$ est remplacée par une fonction continue *non négative* $\psi(x)$ et où, par suite, $\Phi(y)$ est remplacée par la fonction $\Psi(z)$ des probabilités totales de la variable aléatoire $\mathrm{Z} = \psi(\mathrm{X})$.

Supposons d'abord que $\overline{\psi(\mathrm{X})}$ existe au sens que nous avons adopté, c'est-à-dire que l'intégrale

$$\overline{\mathrm{Z}} = \int_{-\infty}^{+\infty} z\, d\Psi(z)$$

soit absolument convergente. Puisque $\psi(\mathrm{X})$ ne peut être négatif, $\Psi(z)$ est nul pour $z \leqq 0$. L'intégrale ci-dessus se réduit donc à

$$\int_{0}^{+\infty} z\, d\Psi(z)$$

Ceci étant, comparons cette intégrale à l'intégrale

$$\int_{a}^{b} \psi(x)\, d\mathrm{G}(x)$$

Soient $z_0 = 0$, z_1, z_2, ... une suite des nombres $\geqq 0$, croissants, tendant vers l'infini et dont les différences $z_i - z_{i-1}$ ont une même borne supérieure δ. En prenant k assez grand, on pourra trouver un z_k supérieur à toutes les valeurs de la fonction continue $\psi(x)$ dans (a, b). Nous pouvons maintenant montrer que

$$\int_{a}^{b} \psi(x)\, d\mathrm{G}(x) \leqq \delta + \int_{0}^{+\infty} z\, d\Psi(z).$$

Soit, en effet, E_i l'ensemble des points x de (a, b) où $\psi(x) < z_i$; ψ étant continu, cet ensemble est formé des points intérieurs à un ensemble F_i de segments ne chevauchant pas. En outre, F_k remplit (a, b).

Donc

$$\int_a^b \psi(x)\,d\,C(x) = \int_{F_l} \psi(x)\,d\,C(x) = \int_{F_1} + \left[\int_{F_2} - \int_{F_1}\right] + \ldots + \left[\int_{F_l} - \int_{F_{l-1}}\right]$$

Or on a

$$0 \leqq \int_{F_i} \psi(x)\,d\,C(x) - \int_{F_{i-1}} \psi(x)\,d\,C(x) \leqq z_i \left[\int_{F_i} d\,C(x) - \int_{F_{i-1}} d\,C(x)\right].$$

Appelons $\Psi_0(z)$ la probabilité que $\psi(X) < z$ *quand* $a \leqq X \leqq b$ On a évidemment

$$\Psi_0(z_i - \varepsilon) \leqq \int_{F_i} d\,C(x) \leqq \Psi_0(z_i + \varepsilon) \qquad \text{pour } \varepsilon > 0$$

Si l'on a soin de prendre pour z_i des points où $\Psi_0(z)$ est continu, ce qui n'est pas contradictoire avec les hypothèses déjà faites sur les z_i, on voit, en faisant tendre ε vers zéro, qu'on aura

$$\int_{F_i} d\,C(x) = \Psi_0(z_i)$$

Par suite

$$\begin{aligned}\int_a^b & \psi(x)\,d\,C(x) \\ & \leqq z_1 \Psi_0(z_1) + z_2[\Psi_0(z_2) - \Psi_0(z_1)] + \ldots + z_k[\Psi_0(z_k) - \Psi_0(z_{k-1})] \\ & \leqq \delta + z_1[\Psi_0(z_2) - \Psi_0(z_1)] + \ldots + z_{k-1}[\Psi_0(z_k) - \Psi_0(z_{k-1})]\end{aligned}$$

Or $\Psi_0(z_i) - \Psi_0(z_{i-1})$ est la probabilité que $z_{i-1} \leqq \psi(X) < z_i$, *quand* X est dans (a, b); cette différence est donc au plus égale à

$$\Psi(z_i) - \Psi(z_{i-1}).$$

Donc

$$\begin{aligned}\int_a^b \psi(x)\,d\,C(x) & \leqq \delta + z_1[\Psi(z_2) - \Psi(z_1)] + \ldots + z_{k-1}[\Psi(z_k) - \Psi(z_{k-1})] \\ & \leqq \delta + \int_{z_1}^{z_k} z\,d\Psi(z) \\ & \leqq \delta + \int_{z_1}^{+\infty} z\,d\Psi(z) \\ & \leqq \delta + \int_0^{+\infty} z\,d\Psi(z)\end{aligned}$$

Ainsi $\int_a^b \Psi(x)\,dC(x)$ est inférieur à un nombre fixe indépendant de a et de b [1]. Par suite, l'élément différentiel étant ≥ 0, l'existence de $\int_{-\infty}^{+\infty} \psi(x)\,dC(x)$ est établie. Et en faisant tendre a et b vers $-\infty$ et $+\infty$ dans l'inégalité précédente, on a

$$\int_{-\infty}^{+\infty} \psi(x)\,dC(x) \leqq \delta + \int_0^{+\infty} z\,d\Psi(z).$$

D'où, en faisant tendre δ vers zéro,

$$(46)\qquad \int_{-\infty}^{+\infty} \psi(x)\,dC(x) \leqq \int_0^{+\infty} z\,d\Psi(z).$$

Inversement, supposons que l'intégrale $\int_{-\infty}^{+\infty} \psi(x)\,dC(x)$ soit finie. On aura, avec les notations précédentes, pour tout entier j

$$\begin{aligned}\int_0^{z_j} z\,d\Psi(z) &= \int_0^{z_1} + \int_{z_1}^{z_2} + \ldots + \int_{z_{j-1}}^{z_j} \\ &\leqq z_1\Psi(z_1) + z_2[\Psi(z_2) - \Psi(z_1)] + \ldots + z_j[\Psi(z_j) - \Psi(z_{j-1})] \\ &\leqq \delta + z_1[\Psi(z_2) - \Psi(z_1)] + \ldots + z_{j-1}[\Psi(z_j) - \Psi(z_{j-1})]\end{aligned}$$

Or, désignons par H_i l'ensemble des points de la droite illimitée $x'x$ où $\psi(x) < z_i$, c'est l'ensemble des points x intérieurs à un ensemble L_i de segments ne chevauchant pas. On a

$$\Psi(z_i - \varepsilon) \leqq \int_{L_i} dC(x) \leqq \Psi(z_i + \varepsilon)$$

pour tout $\varepsilon > 0$. En ayant soin de prendre les z_i distincts des points de discontinuité de $\Psi(z)$, on aura

$$\int_{L_i} dC(x) = \Psi(z_i)$$

---

[1] Il suffit que ceci soit démontré pour une infinité dénombrable de valeurs de a et de b, ce qui permettra de prendre les z_i distincts des points de discontinuité de tous les Ψ_0 correspondants.

D'où

$$
\begin{aligned}
\int_0^{z_j} z\,d\Psi(z) &\leqq \delta + z_1\left[\int_{I_2} dC(x) - \int_{I_1} dC(x)\right] \\
&\quad + \dots \\
&\quad + z_{j-1}\left[\int_{I_j} dC(x) - \int_{I_{j-1}} dC(x)\right] \\
&\leqq \delta + \left[\int_{I_2} \psi(x)\,dC(x) - \int_{I_1} \psi(x)\,dC(x)\right] \\
&\quad + \dots \\
&\quad + \left[\int_{I_j} \psi(x)\,dC(x) - \int_{I_{j-1}} \psi(x)\,dC(x)\right] \\
&= \delta + \int_{I_j} \psi(x)\,dC(x) - \int_{I_1} \psi(x)\,dC(x) \\
&\leqq \delta + \int_0^{+\infty} \psi(x)\,dC(x) - \int_{I_1} \psi(x)\,dC(x).
\end{aligned}
$$

Alors, quel que soit j, l'intégrale $\int_0^{z_j} z\,d\Psi(z)$ reste inférieure à un nombre fixe indépendant de j. Par conséquent, l'intégrale $\int_0^{+\infty} z\,d\Psi(z)$ est finie, et l'on a d'après l'inégalité précédente

$$\bar{Z} = \int_0^{+\infty} z\,d\Psi(z) \leqq \delta + \int_0^{+\infty} \psi(x)\,dC(x)$$

et comme δ est un nombre positif arbitraire

$$\bar{Z} \leqq \int_0^{+\infty} \psi(x)\,dC(x). \tag{46 bis}$$

D'ailleurs, en combinant les inégalités (46) et (46^{bis}), on voit qu'on a

$$\bar{Z} = \int_0^{+\infty} \psi(x)\,dC(x). \tag{47}$$

En résumé, si $\psi(x)$ est une fonction continue $\geqq 0$ et si $C(x)$ est la fonction des probabilités totales de X, la condition nécessaire et suffisante pour que $\overline{\psi(X)}$ existe est que $\int_0^{+\infty} \psi(x)\,dC(x)$ soit une intégrale convergente et alors

$$\overline{\psi(X)} = \int_0^{+\infty} \psi(x)\,dC(x).$$

II. Passons maintenant au cas d'une fonction $\varphi(x)$, encore continue, mais de signe quelconque.

Considérons maintenant les deux fonctions

$$\psi(x) = |\varphi(x)| + \varphi(x), \qquad \theta(x) = |\varphi(x)| - \varphi(x)$$

Si $\varphi(x)$ est continue, $\psi(x)$ et $\theta(x)$ seront deux fonctions continues $\geqq$ o. Nous savons que si l'intégrale

$$\int_{-\infty}^{+\infty} y\, d\Phi(y)$$

est absolument convergente $\overline{\varphi(x)}$ et $\overline{|\varphi(x)|}$ existent, donc aussi les valeurs moyennes de $\psi(x)$ et $\theta(x)$. Ayant prouvé la proposition en vue, dans le cas de fonctions continues $\geqq$ o, nous avons donc établi que

$$\int_{-\infty}^{+\infty} \psi(x)\, d\,C(x) \quad \text{et} \quad \int_{-\infty}^{+\infty} \theta(x)\, d\,C(x)$$

sont convergentes. Donc leur demi-somme

$$\int_{-\infty}^{+\infty} |\varphi(x)|\, d\,C(x)$$

sera convergente et par suite

$$\int_{-\infty}^{+\infty} \varphi(x)\, d\,C(x)$$

sera absolument convergente. De plus, on aura

$$\overline{|\varphi(x)|} + \overline{\varphi(x)} = \overline{\psi(x)} = \int_{-\infty}^{+\infty} \psi(x)\, d\,C(x),$$

$$\overline{|\varphi(x)|} - \overline{\varphi(x)} = \overline{\theta(x)} = \int_{-\infty}^{+\infty} \theta(x)\, d\,C(x),$$

d'où

$$\overline{\varphi(x)} = \int_{-\infty}^{+\infty} \varphi(x)\, d\,C(x).$$

Réciproquement, si

$$\int_{-\infty}^{+\infty} \varphi(x)\, d\,C(x)$$

est absolument convergente, les intégrales

$$\int_{-\infty}^{+\infty} \frac{[|\varphi(x)| + \varphi(x)]}{2} d\,C(x),$$

$$\int_{-\infty}^{+\infty} \frac{[|\varphi(x)| - \varphi(x)]}{2} d\,C(x)$$

seront convergentes. Comme les crochets sont des fonctions $\frac{\psi(x)}{2}$, $\frac{\theta(x)}{2}$ qui sont continues et ≥ 0, les valeurs moyennes $\left(\overline{\frac{\psi(X)}{2}}\right)$ et $\left(\overline{\frac{\theta(X)}{2}}\right)$ existent et sont égales à ces intégrales. Par suite, la valeur moyenne de $\frac{\psi(X)}{2} - \frac{\theta(X)}{2}$, c'est-à-dire de $\varphi(X)$ existe et est égale à la différence de ces intégrales, c'est-à-dire à $\int_{-\infty}^{+\infty} \varphi(x)\,d\,C(x)$.

En résumé, *la condition nécessaire et suffisante pour que la valeur moyenne $\overline{\varphi(X)}$ d'une fonction continue d'une variable aléatoire X existe est que l'intégrale*

$$\int_{-\infty}^{+\infty} \varphi(x)\,d\,C(x), \tag{48}$$

où $C(x)$ est la fonction des probabilités totales de X, non seulement existe mais soit une intégrale absolument convergente. Et alors, on a

$$\overline{\varphi(X)} = \int_{-\infty}^{+\infty} \varphi(x)\,d\,C(x). \tag{49}$$

D'ailleurs, la définition de $\overline{\varphi(x)}$ par la formule (44) peut garder un sens lorsque φ est discontinue. Elle donne, par exemple, une valeur bien déterminée si la fonction $\varphi(x)$ est bornée et $\Phi(x)$, bien déterminée. Il peut être utile d'étendre de même l'égalité (47) à des cas plus généraux que celui où $\varphi(x)$ est continue. M. Glivenko (2) a prouvé ([1]) que la condition nécessaire et suffisante ci-dessus s'étend au cas où $\varphi(x)$ est une fonction mesurable au sens de M. Borel.

Il nous suffira pour la suite d'observer que la démonstration

---

([1]) La démonstration, très abrégée, de cette note utilise des définitions et propriétés — que nous n'avons pas voulu supposer connues — de l'intégrale de Stieltjes dans des cas très généraux.

précédente et, par suite, la formule (49) s'étendent au cas où $\varphi(x)$ est une fonction monotone, continue ou non à droite, mais continue à gauche. Or nous avons montré ailleurs (Fréchet, 183) que la définition de l'intégrale de Stieltjes

$$\int_{-\infty}^{+\infty} \varphi(x)\,dv(x),$$

originairement donnée pour $\varphi(x)$ continue [et $v(x)$ à variation bornée], s'étend au cas où $\varphi(x)$, continue ou non, est, comme $v(x)$, à variation bornée, sous les conditions suivantes : 1° les fonctions $\varphi(x)$, $v(x)$ ne doivent nulle part être discontinues de deux côtés opposés; 2° la sommation

$$\Sigma\, \varphi(\xi_i)\,[v(x_i) - v(x_{i-1})],$$

qui donne à la limite la valeur de l'intégrale, étant effectuée pour une suite de divisions D_n de la droite par des points ..., x_{i-1}, x_i, ..., on devra choisir cette suite de sorte que les points de discontinuité communs à φ et à v appartiennent chacun aux D_n à partir d'un certain rang (éventuellement variable avec le point).

Cette dernière condition n'intervient pas dans la démonstration donnée plus haut. Et puisque toute fonction de probabilité totale est continue à gauche, on voit que la première condition exprime que la démonstration sera valable précisément dans le cas où elle sera utile, c'est-à-dire où l'intégrale de Stieltjes a un sens.

Remarque. — Chemin faisant, rappelons que deux expressions principales de la moyenne de $\varphi(X)$ se présentent dans les applications, l'une de la forme

$$\sum p_i\, \varphi(x_i),$$

l'autre de la forme

$$\int \varphi(x)\,\theta(x)\,dx.$$

La première se présente quand la variable aléatoire X ne prend qu'un nombre fini (ou un ensemble dénombrable) de valeurs x_i; la seconde quand il y a une densité de probabilité $\theta(x)$. L'intégrale de Stieltjes a le premier avantage de représenter l'une ou l'autre forme indistinc-

tement par le même symbole

$$\int \varphi(x)\, d\,\mathrm{C}(x)$$

Mais elle a aussi ce second avantage de permettre également la représentation d'une troisième forme irréductible aux deux premières.

Nous avons, en effet, fait observer ailleurs (Fréchet, 47) (et ce résultat a été ensuite généralisé par MM de La Vallée-Poussin et Lebesgue) que toute intégrale de Stieltjes

$$\int_a^b \varphi(x)\, d\,v(x),$$

où φ est continue et v à variation bornée peut être décomposée en 3 parties une somme ou série de la forme

$$\sum_j p_j\, \varphi(x_j),$$

où les p_j sont les sauts de $v(x)$, une intégrale de Lebesgue de la forme

$$\int_a^b \varphi(x)\,\theta(x)\, dx$$

et une troisième partie

$$\int_a^b \varphi(x)\, d\,\alpha(x)$$

qui est une intégrale de Stieltjes d'un type *non réductible* aux deux précédents et où $\alpha(x)$ est une fonction continue à variation bornée *indépendante* de φ et dont la dérivée est nulle presque partout Cette troisième partie est ce qui constitue la véritable nouveauté de l'intégrale de Stieltjes, car elle est d'un type mixte, formation intermédiaire entre la série et l'intégrale de Lebesgue

Si l'emploi de l'intégrale de Stieltjes est indiqué quand on veut faire un raisonnement applicable *simultanément* aux deux premiers types, il faut reconnaître que ces types sont au contraire, quand on veut les traiter séparément, souvent plus maniables sous leurs formes primitives soit de série, soit d'intégrale de Lebesgue Il n'en est pas de même pour la troisième partie pour laquelle la forme de l'intégrale de Stieltjes semble plus appropriée que toute autre connue. Toutefois, nous avons fait observer ailleurs (Fréchet, 47) que l'on pouvait peut-être avec avantage calculer l'intégrale

$$\mathrm{J} = \int_a^b \varphi(x)\, d\,\alpha(x)$$

au moyen de la formule

$$\int_a^b \varphi(x)\, d\alpha(x) = \lim_{(\text{mes } S_E \to 0)} \sum_{S_E} \varphi(\xi_n)\, \Delta_{I_n} \alpha(x),$$

où S_E est un ensemble d'intervalles I_n couvrant l'ensemble E des points où la dérivée de $\alpha(x)$ est infinie, où ξ_n est un point de E choisi arbitrairement dans I_n, et où enfin quand I_n est l'intervalle x'_n, x''_n, on a posé

$$\Delta_{I_n} \alpha(x) = \alpha(x''_n) - \alpha(x'_n)$$

Contrairement à ce que pourrait laisser supposer la définition directe de J comme intégrale de Stieltjes, on voit par cette formule que cette intégrale de Stieltjes particulière ne dépend que des valeurs de $\varphi(x)$ sur E, c'est-à-dire sur un ensemble de mesure nulle indépendant de $\varphi(x)$

Nous avons donné (47) comme exemple le cas où $\alpha(x)$ est une fonction en escalier considérée par Cantor, constante sur des intervalles disjoints dont la longueur totale est égale à celle de l'intervalle d'intégration, cas où l'on a

$$\int_a^b \varphi(x)\, d\alpha(x) = \lim_{n \to \infty} \frac{f(\xi_1^{(n)}) + \ldots + f(\xi_{2^n}^{(n)})}{2^n},$$

les $\xi_p^{(n)}$ étant pris convenablement sur un certain ensemble de mesure nulle

Moments. — On appelle moment d'ordre r d'une variable aléatoire X, la valeur moyenne de X^r, r étant un nombre certain supposé positif. X^r a une signification bien précise quand r est un entier. Quand r n'est pas entier, on se bornera, pour éviter toute difficulté aux moments des variables aléatoires X qui ne sont jamais négatives (*voir* cependant, p. 65).

Comme x^r est une fonction continue, la condition nécessaire et suffisante pour que le moment d'ordre r de X existe est, d'après ce qui précède, que l'intégrale $\int_{-\infty}^{+\infty} x^r\, d\,C(x)$ existe et soit absolument convergente. Si cette condition est remplie, on aura

$$\overline{X^r} = \int_{-\infty}^{+\infty} x^r\, d\,C(x). \tag{50}$$

Écarts moyens. — Quel que soit le nombre positif r, on appelle écart moyen d'ordre r de deux variables aléatoires X, Y, la racine d'ordre r de la valeur moyenne de $|X - Y|^r$.

Si

$$q(t) = \text{Pr.}\{|X - Y| < t\},$$

on voit que la condition nécessaire et suffisante pour que cet écart, $\mu^{(r)}$, existe est que l'intégrale $\int_0^{+\infty} t^r \, dq(t)$ soit absolument convergente, et alors

$$\mu^{(r)} = \sqrt[r]{\int_0^{+\infty} t^r \, dq(t)} \tag{51}$$

Puisque, pour $t > 1$, t^r croît avec r, il est clair que pour chaque couple de variables aléatoires X, Y, leurs écarts moyens, ou bien existent pour tous les ordres, ou bien sont infinis pour tous les ordres, ou bien existent pour r inférieur à un certain nombre ρ et sont infinis pour r supérieur à ρ. On peut d'ailleurs préciser. Soit α un nombre arbitraire; on a, pour r et s positifs,

$$0 \leqq \int_0^{+\infty} \left[\alpha \,|\, t\,|^{\frac{r}{2}} + |\, t\,|^{\frac{s}{2}}\right]^2 dq(t).$$

D'où, en posant $u = \frac{r+s}{2}$, et en appelant $\nu^{(r)}$ le moment d'ordre r de $|\, X - Y\,|$:

$$0 \leqq \alpha^2 \nu^{(r)} + 2\alpha\nu^{(u)} + \nu^{(s)}$$

et, par suite,

$$\nu^{(u)} \leqq \sqrt{\nu^{(r)} \nu^{(s)}} \tag{52}$$

ou

$$\log \nu^{\left(\frac{r+s}{2}\right)} \leqq \frac{1}{2} [\log \nu^{(r)} + \log \nu^{(s)}]$$

D'ailleurs, il ne pourrait y avoir égalité que s'il existait une valeur α' de α telle que

$$\int_0^{+\infty} \left[\alpha' \,|\, t\,|^{\frac{r}{2}} + |\, t\,|^{\frac{s}{2}}\right]^2 dq(t) = 0,$$

c'est-à-dire si

$$\left[\alpha' \,|\, X - Y\,|^{\frac{r}{2}} + |\, X - Y\,|^{\frac{s}{2}}\right]^2$$

avait une valeur moyenne nulle. Ceci ne pourrait avoir lieu, puisque cette quantité est $\geqq 0$, que si la probabilité pour qu'elle soit $\neq 0$, était nulle.

Or, elle ne peut être nulle que si $X = Y$ ou si $|\, X - Y\,|$ est égale à

une valeur certaine $A = [-\alpha']^{\frac{2}{s-r}}$ qui peut être $\neq 0$. Donc, le cas d'exception ne peut se présenter que si $|X - Y|$ n'a au plus qu'une valeur possible différente de zéro. Ou, plus précisément, si $|X - Y|$ est « presque certainement » égal soit à zéro, soit à une valeur certaine $A \neq 0$. Dans ce cas exceptionnel, il est clair que si p_1 est la probabilité que $X - Y = A$, p_2 la probabilité que $X - Y = -A$ et $p = p_1 + p_2$, on aura

$$\nu^{(r)} = A^r p, \qquad \text{d'où} \qquad \nu^{(u)} = \sqrt{\nu^{(r)}\nu^{(s)}}.$$

En dehors de ce cas exceptionnel, on a

$$\log \nu^{\left(\frac{r+s}{2}\right)} < \frac{1}{2}[\log \nu^{(r)} + \log \nu^{(s)}],$$

la courbe représentative de la fonction de r, $y = \log \nu^{(r)}$ tourne donc sa concavité vers les y positifs. Elle passe par l'origine puisque $\nu^{(0)} = 1$. Dès lors, le coefficient angulaire

$$\frac{1}{r} \log \nu^{(r)} = \log \mu^{(r)}$$

de la droite qui joint un point d'abscisse r de cette courbe à l'origine est une fonction croissante de r. Il en est donc de même de $\mu^{(r)}$. Ainsi, *l'écart moyen, μ_r, d'ordre positif r, de X et de Y, tant qu'il reste fini, croît constamment et effectivement avec cet ordre r.*

Dans le cas exceptionnel où $|X - Y| = 0$ ou $A > 0$ avec des probabilités respectives q et p, dont la somme est égale à l'unité, on a $\mu^{(r)} = A p^{\frac{1}{r}}$ quel que soit $r > 0$ et, par suite, $\mu^{(r)}$ croît encore avec r pourvu que p soit différent de 1 et de 0. Si $p = 1$ ou 0, on est dans le cas où $|X - Y|$ est constant ou presque certainement constant : $\mu^{(r)}$ est fini et constant quand r croît [1].

Signalons en passant une inégalité obtenue par M. Slutsky après un calcul assez laborieux et qui résulte immédiatement de l'inégalité

---

[1] Il ne serait donc pas exact de dire que cette constance de $\mu^{(r)}$ quand r croît ne peut avoir lieu que si $X - Y$ est presque certainement égal à une valeur certaine, puisque $X - Y$ peut prendre *deux* valeurs distinctes A, $-A$, avec des probabilités toutes deux > 0 et < 1 (mais de somme égale à l'unité).

évidente

$$\alpha^{r+s} + \beta^{r+s} - \alpha^r \beta^s \geqq 0$$

où α, β, r, s sont des nombres certains $\geqq 0$. Si v, w sont des nombres certains arbitraires et X, Y des variables aléatoires définies sur la même catégorie d'épreuves, on aura pour chaque épreuve

$$|X - v|^{r+s} + |Y - w|^{r+s} - |X - v|^r \, |Y - w|^s \geqq 0$$

La moyenne du premier membre sera, par suite, $\geqq 0$; d'ou l'inégalité de Slutsky (1, p. 32)

$$\mathfrak{M}\{|X - v|^r \, |Y - w|^s\} \leqq \mathfrak{M}|X - v|^{r+s} + \mathfrak{M}|Y - w|^{r+s}$$

Quand X et Y sont indépendants, elle devient $M_r N_s \leqq M_{r+s} + N_{r+s}$, en appelant M_r, N_r les moments d'ordre r de $|X - v|$ et $|Y - w|$.

On peut d'ailleurs préciser l'inégalité de Slutsky sous la forme

$$(52^a) \qquad \mathfrak{M}|X|^r |Y|^s \leqq k(r, s)\{\mathfrak{M}|X|^{r+s} + \mathfrak{M}|Y|^{r+s}\}$$

ou $k(r, s)$ désigne le maximum de $\dfrac{\alpha^r \beta^s}{\alpha^{r+s} + \beta^{r+s}}$ quand α et β y prennent toutes les valeurs positives. On vérifie facilement qu'on a toujours $k(r, s) < 1$ et même qu'on a

$$(52^b) \qquad k(r, s) = \left(\frac{s}{r+s}\right)^{\frac{s}{r+s}} \left(\frac{r}{r+s}\right)^{\frac{r}{r+s}}$$

Cette quantité est non seulement inférieure à la valeur 1 par laquelle elle est remplacée dans l'inégalité de Slutsky, mais encore elle est au plus égale à $\frac{1}{2}$. C'est là sa plus grande valeur, atteinte pour $r = s$

En particulier, si $r = s = 1$, on obtient l'inegalité

$$(52^c) \qquad \mathfrak{M}|XY| \leqq \frac{1}{2}\{\mathfrak{M}X^2 + \mathfrak{M}Y^2\}$$

dont la démonstration directe est évidente et qui est, d'ailleurs, plus faible que l'inégalité (56^c) (de la p. 70) dont elle peut se déduire.

On appelle *écart quadratique moyen* l'écart d'ordre deux et *écart moyen* tout court, l'écart d'ordre un : ce sont les deux plus utiles. Nous voyons que l'*écart moyen σ de deux variables aléatoires est toujours au plus égal à leur écart quadratique moyen μ.*

On a même $\sigma < \mu$ si la valeur absolue de la différence des deux variables n'est pas presque certainement constante.

On obtient d'ailleurs facilement une interprétation de la différence $\mu^2 - \sigma^2$ en observant qu'elle est égale à

$$
(53) \qquad \mathfrak{M}[(X-Y)^2] - [\mathfrak{M}(|X-Y|)]^2 = \frac{1}{2}\int_0^{+\infty}\left\{\int_0^{+\infty}(t-t_1)^2\,dq(t_1)\right\}dq(t).
$$

En effet,

$$
\begin{aligned}
\mu^2-\sigma^2 &= \int_0^{+\infty} t^2\,dq(t) - \left[\int_0^{+\infty} t_1\,dq(t_1)\right]^2 \\
&= \left[\int_0^{+\infty} dq(t_1)\right]\left[\int_0^{+\infty} t^2\,dq(t)\right] \\
&\quad - \left[\int_0^{+\infty} t_1\,dq(t_1)\right]\left[\int_0^{+\infty} t\,dq(t)\right],
\end{aligned}
$$

ce à quoi se réduit le second membre de (53) après développement.

Inégalité de Gauss. — D'après les résultats obtenus plus haut, on a, en particulier,

$$
\frac{\mu_1}{\mu_2} \geqq 1
$$

Gauss avait perfectionné cette inégalité *dans le cas où :* 1° il existe, pour la déviation absolue de la variable aléatoire considérée, une densité de probabilité ; 2° celle-ci est non croissante. Il avait indiqué, sans démonstration (Gauss, 1, Art. 10), que l'on a, dans ce cas,

$$
\frac{\mu_1}{\mu_2} \geqq \sqrt[4]{\frac{9}{5}}.
$$

Cette inégalité a été généralisée en 1866 par Winckler (1), et dans la même hypothèse, sous la forme

$$
\frac{\mu^{(p)}}{\mu^{(n)}} \geqq \frac{(n+1)^{\frac{1}{n}}}{(p+1)^{\frac{1}{p}}} \qquad \text{pour } p > n,
$$

formule également plus précise, mais d'une validité plus restreinte que la formule $\frac{\mu^{(p)}}{\mu^{(n)}} \gtreqless 1$ obtenue plus haut.

L'inégalité de Gauss a été démontrée par Krüger en 1896, et celle

de Winckler par G. Faber (1) en 1922. Nous donnerons ici une démonstration plus simple, publiée en 1931 par M. von Mises (2, p. 187).

Nous avons ici

$$\nu^{(n)} = \int_0^{+\infty} t^n \, d\varphi(t) = \int_0^{+\infty} \alpha(t)\, t^n \, dt,$$

où $\alpha(t)$ est une fonction non croissante. Or, cette intégrale peut s'écrire

$$(54) \qquad \left[\frac{t^{n+1}}{n+1}\alpha(t)\right]_0^{+\infty} - \frac{1}{n+1}\int_0^{+\infty} t^{n+1}\, d\alpha(t).$$

Supposons d'abord $\alpha(o)$ fini; alors en posant

$$Q(t) = 1 - \frac{\alpha(t)}{\alpha(o)},$$

$Q(t)$ sera une fonction non décroissante variant de o à 1. Car on a

$$1 = \int_0^{+\infty} d\varphi(t) = \int_0^{+\infty} \alpha(t)\, dt,$$

ce qui ne peut avoir lieu pour α non croissant que si $\alpha(t)$ tend vers zéro avec $\frac{1}{t}$. Dès lors,

$$-\frac{1}{n+1}\int_0^{+\infty} t^{n+1}\, d\alpha(t) = \frac{\alpha(o)}{n+1}\rho^{(n+1)},$$

où $\rho^{(n+1)}$ est le moment d'ordre $n+1$ correspondant à la fonction des probabilités totales $Q(t)$. D'autre part, le terme $\left[\frac{t^{n+1}\alpha(t)}{n+1}\right]_0^{+\infty}$ est nul. Car $t^{n+1}\alpha(t)$ tend évidemment vers zéro quand $t \to o$ puisque $\alpha(o)$ est fini et quand, au contraire, $t \to +\infty$, $t^{n+1}\alpha(t)$ tend encore vers zéro [1].

---

[1] En effet, on a, pour a positif arbitraire,

$$\int_{\frac{a}{2}}^{+\infty} t^n \alpha(t)\, dt \geqq \int_{\frac{a}{2}}^{a} t^n \alpha(t)\, dt \geqq \frac{a^n}{2^n}\alpha(a)\frac{a}{2};$$

d'où

$$o \leqq a^{n+1}\alpha(a) \leqq 2^{n+1}\int_{\frac{a}{2}}^{+\infty} t^n \alpha(t)\, dt.$$

Si donc $\nu^{(n)}$, c'est-à-dire $\int_0^{+\infty} t^n \alpha(t) dt$, est fini, le dernier membre, donc aussi le second, tendent vers zéro quand $a \to +\infty$

On a donc enfin

$$\nu^{(n)} = \frac{\alpha(0)}{n+1}\,\rho^{(n+1)}$$

Or, on a vu que pour le moment $\rho^{(n+1)}$, on a

$$\log\rho^{(n+1)} = \log\rho^{\left(\frac{r+1+s+1}{2}\right)} \leqq \frac{1}{2}\,[\log\rho^{(r+1)} + \log\rho^{(s+1)}],$$

d'où

$$\log\delta^{\left(\frac{r+s}{2}\right)} \leqq \frac{1}{2}\,[\log\delta^{(r)} + \log\delta^{(s)}]$$

en posant

$$\delta^{(n)} = (n+1)\,\nu^{(n)}.$$

Dès lors, la courbe $y = \log\delta^{(x)}$ est concave vers les $y > 0$, et comme elle passe par l'origine, on voit encore que $\frac{1}{r}\log\delta^{(r)}$ est une fonction non décroissante de r. Donc, pour $p > n$,

$$[(p+1)\nu^{(p)}]^{\frac{1}{p}} \geqq [(n+1)\nu^{(n)}]^{\frac{1}{n}}$$

Ainsi, *quand* la densite de probabilité de $(X - Y)$, $\sigma(t)$, existe et est non décroissante, ce n'est plus seulement l'écart moyen $\mu^{(r)}$ d'ordre r de $|X - Y|$, *c'est aussi* $(r+1)^{\frac{1}{r}}\mu^{(r)}$ *qui est une fonction non décroissante de* r

Toutefois, pour aller plus vite au but, nous avons laissé de côté le cas où $\sigma(0)$ est infini.

Dans ce cas, si $\nu^{(n)}$ est fini, $\nu^{(k-1)}$ est fini pour $k \leq n+1$, et alors (¹) l'intégrale

$$J = \int_0^{+\infty} t^k\, d[-\alpha(t)]$$

(¹) On a

$$\int_0^{\varepsilon} t^n \alpha(t)\,dt \geqq \alpha(\varepsilon)\int_0^{\varepsilon} t^n\,dt = \frac{\varepsilon^{n+1}\alpha(\varepsilon)}{n+1} \geqq 0$$

Puisque $\nu^{(n)}$ est fini, $\int_0^{\varepsilon} t^n \alpha(t)\,dt$ tend vers zéro avec ε. Donc aussi $\varepsilon^{n+1}\alpha(\varepsilon)$.
On a vu aussi que $a^{n+1}\alpha(a)$ tend vers zéro avec $\frac{1}{a}$. Dès lors, dans l'expression (54), $\left[\frac{t^{n+1}\alpha(t)}{n+1}\right]_0^{+\infty}$ est toujours nul, et si $\nu^{(n)}$ est fini, il en est de même de

$$\int_0^{+\infty} t^{n+1}\,d\alpha(t)$$

est finie. Alors on posera

$$Q_1(t) = \frac{\int_0^t u^k d[-\alpha(u)]}{J};$$

d'où

$$dQ_1(t) = -\frac{t^k d\alpha(t)}{J}$$

et si $\nu^{(n)}$ est fini,

$$-\frac{1}{n+1}\int_0^{+\infty} t^{n+1} d\alpha(t) = \frac{J}{n+1}\int_0^{+\infty} t^{n+1-k} dQ_1(t) = \frac{J}{n+1} l^{(n+1-k)},$$

en appelant $l^{(n+1-k)}$ le moment d'ordre $n+1-k$ ($\geqq 0$) correspondant à la fonction des probabilités totales $Q_1(t)$. On aura alors

$$\nu^{(n)} = \frac{J}{n+1} l^{(n+1-k)},$$

d'où encore

$$\log \delta^{\left(\frac{r+s}{2}\right)} \leqq \frac{1}{2}[\log \delta^{(r)} + \log \delta^{(s)}],$$

et l'on arrive à la même conclusion.

Moments algébriques. — Quand r est un entier positif, $(X-Y)^r$ a une signification et peut se distinguer de $|X-Y|^r$ quand r est un impair

Pour étendre cette expression au cas où r n'est pas entier, M. Slutsky considère les deux expressions

$$|X-Y|^r \quad \text{et} \quad [\operatorname{sgn}(X-Y)]\,|X-Y|^r,$$

où $\operatorname{sgn}(X-Y) = -1$ si $X-Y \leqq 0$, $=+1$ si $X-Y > 0$. Elles coïncident respectivement avec $(X-Y)^r$, la première quand r est entier pair, la seconde quand r est entier impair. On pourra désigner par M_r et m_r leurs valeurs moyennes respectives. M. Slutsky (1, p 36) en déduit un curieux moyen de définir la médiane et la moyenne de façons analogues. La valeur moyenne de X est racine de l'équation en v, $m_1 = 0$ quand Y est un nombre certain v. D'autre part, on a

$$\begin{aligned} m_0 &= \mathfrak{M}\{[\operatorname{sgn}(X-v)]\,|X-v|^0\} = \mathfrak{M}\{\operatorname{sgn}(X-v)\} \\ &= -\int_{-\infty}^{v} dC(x) + \int_{v}^{+\infty} dC(x) \\ &= 1 - 2C(v). \end{aligned}$$

Quand il existe un nombre α tel que $C(\alpha)=\frac{1}{2}$, on voit que la médiane est racine de l'équation $m_0=0$. Dans le cas général, on a vu, p. 39, que c'est un nombre α tel que $1-C(v+0)-C(v-0)$ soit $\geqq 0$ pour $v<\alpha$ et $\leqq 0$ pour $v>\alpha$, et par suite tel que

$$C(\alpha-0)\leqq\frac{1}{2}\leqq C(\alpha+0).$$

Une médiane est donc un nombre α tel que $m_0(v)$ s'annule ou au moins change de signe quand v passe en croissant par la valeur α.

Remarque. — Nous généraliserons plus loin (p. 120), la notion d'écart moyen d'ordre r. Nous établirons à ce moment quelques-unes des propriétés des écarts moyens (p. 120), propriétés qui justifient le nom adopté et qui sont l'origine de l'introduction et de l'emploi répété de ces écarts en Statistique aussi bien qu'en Théorie des probabilités.

Écart médian. — Signalons aussi qu'on fait souvent usage de l'écart médian, dit aussi écart probable. Nous appellerons *écart médian* de deux variables aléatoires X, Y la valeur médiane de la valeur absolue de leur différence. C'est donc, en gros, un nombre m, tel qu'il y ait des probabilités égales pour que $|X-Y|>m$, ou pour que $|X-Y|<m$. Il peut d'ailleurs n'y avoir dans certains cas aucun tel nombre si la probabilité

$$q(t)=\{\text{Pr.}[|X-Y|<t]\}$$

est discontinue. On peut, dans ce cas, comme on l'a vu plus haut, adopter une définition plus générale, mais moins intuitive. L'écard médian m sera un nombre tel que les probabilités, pour que $|X-Y|>m$ et pour que $|X-Y|<m$ soient toutes deux $\leqq\frac{1}{2}$. Il peut d'ailleurs aussi y avoir indétermination pour m sur un « segment médian » à l'intérieur duquel $q(t)$ reste égal à $\frac{1}{2}$.

Valeur moyenne du produit de deux variables aléatoires indépendantes. — On n'insiste pas toujours assez sur la différence des champs d'application des formules classiques donnant les valeurs moyennes de $X+Y$ et de XY connaissant celles de X et de Y. Alors

que la formule qu'on a établie plus haut, concernant la moyenne d'une somme, suppose seulement l'existence des moyennes des deux termes de cette somme, on peut citer des exemples simples où la formule qui va nous occuper,

$$\overline{XY} = \overline{X}\,\overline{Y}$$

contient trois moyennes finies et déterminées, et pourtant se trouve en défaut. On peut, au contraire, la démontrer comme suit quand X et Y sont supposés *indépendants*.

I. Dans le cas où X, Y ne peuvent prendre qu'un nombre fini de valeurs distinctes, on a, avec les notations employées précédemment dans ce même cas, p. 41,

$$\overline{XY} = \sum_k p_k u_k = \sum_k \left\{ \sum_{x_i y_j = u_k} p_{i,j} x_i y_j \right\},$$

où $u_k = x_i y_j$. Or, puisque X, Y sont indépendants, $p_{i,j} = p'_i p''_j$. Donc,

$$\overline{XY} = \sum_k \left\{ \sum_{x_i y_j = u_k} p'_i x_i p''_j y_j \right\} = \sum_i p'_i x_i \sum_j p''_j y_j = \overline{X}\,\overline{Y}.$$

II. Dans le cas général, supposons que $\overline{X}$, $\overline{Y}$ existent. On a alors les deux formules (42 *bis*) comme précédemment. D'où

$$\overline{X}.\overline{Y} = \sum_{i,j} \xi_i \eta_j p_{i,j},$$

où le second membre est une série double absolument convergente comme les deux séries simples dont il est le produit.

Or, $p_{i,j}$ est la probabilité pour que le point (X, Y) se trouve dans la région $R_{i,j}$,

$$(55) \qquad x_{i-1} \leqq X < x_i, \qquad y_{j-1} \leqq Y < y_j.$$

Si l'on forme tous les produits $u_k = x_i y_j$, qu'on peut supposer distincts, en prenant, par exemple, $x_i = (i + \sqrt{3})\delta$, $y_j = (j + \sqrt{2})\delta$, et si l'on trace toutes les hyperboles $XY = u_h$, la région R_{ij} se trouvera décomposée en un nombre fini de régions telles que R_{ijh}, où l'on a (55) et

$$u_{h-1} \leqq XY < u_h.$$

Soit p_{ijh} la probabilité que le point (X, Y) soit dans cette région.

Alors

$$p_{ij} = \sum_h p_{ijh} \qquad \text{et} \qquad \overline{XY} = \sum_{i,j,h} \xi_i \eta_j p_{i,j,h}$$

Or,

$$\xi_i \eta_j - u_h = (x_i - \theta\delta)(y_j - \theta'\delta) - (x_i - \theta''\delta)(y_j - \theta'''\delta),$$

les θ, θ', θ'', θ''' étant compris entre o et 1. Donc,

$$|\xi_i \eta_j - u_h| \leqq \delta |y_j| + \delta |x_i| + \delta^2.$$

Alors, la série absolument convergente

$$\overline{XY} = \sum_{i,j,h} \xi_i \eta_j p_{i,j,h}$$

est la somme de deux séries, la série

$$S = \sum_{i,j,h} u_h p_{i,j,h}$$

et la série

$$S' = \sum_{i,j,h} (\xi_i \eta_j - u_h) p_{i,j,h}.$$

Cette dernière est absolument convergente puisqu'elle est majorée par la série

$$S'' = \sum_{i,j,h} [\delta |y_j| + \delta |x_i| + \delta^2] p_{i,j,h}$$

$$= \delta \sum_j |y_j| \sum_{i,h} p_{i,j,h} + \delta \sum_i |x_i| \sum_{j,h} p_{i,j,h} + \delta^2$$

$$= \delta \sum_j |y_j| p''_j + \delta \sum_i |x_i| p'_i + \delta^2$$

et que dans cette expression, les coefficients de δ sont des séries absolument convergentes (car les valeurs moyennes de $|Y|$ et de $|X|$ existent par hypothèse). Donc, la série S est absolument convergente. Lorsque δ tend vers zéro, S'', donc S' tendent vers zéro. Ainsi, S tend vers $\overline{X}.\overline{Y}$. Mais

$$S = \sum_h u_h \sum_{i,j} p_{i,j,h} = \sum_h u_h \{ \text{Pr.}[u_{h-1} \leqq XY < u_h] \}.$$

Par définition, la limite de S est la valeur moyenne de XY, si cette

valeur moyenne existe, et cette existence est garantie par l'absolue convergence de S en vertu de la proposition de la page 38. En résumé,

(56) $$\boxed{\overline{XY} = \overline{X}.\overline{Y} \qquad \textit{quand X et Y sont indépendants}}$$

Et l'existence des moyennes de X et de Y garantit l'existence de la moyenne du produit XY (toujours dans l'hypothèse de l'indépendance de X et de Y).

C'est, en particulier, ce qui a lieu si Y se réduit à un nombre certain a. On a donc

(57) $$\boxed{\overline{aX} = a\overline{X} \qquad \text{quand } a \text{ est un nombre certain.}}$$

Dès lors,

$$\overline{X - Y} = \overline{X + (-1)Y} = \overline{X} + (-1)\overline{Y},$$

d'où

$$\overline{X - Y} = \overline{X} - \overline{Y}.$$

On voit aussi par récurrence que

(57 *bis*) $$\overline{a_1 X_1 + \ldots + a_n X_n} = a_1 \overline{X_1} + \ldots + a_n \overline{X_n}$$

quand $a_1, \ldots, a_n$ sont des nombres certains.

Généralisation. — Revenons au cas I où X, Y ne peuvent prendre qu'un nombre fini de valeurs distinctes, mais ne les supposons plus indépendantes. On aura cette fois $p_{ij} = p'_i p''_{ij}$ où p''_{ij} est la probabilité pour que $Y = y_j$ *quand* $X = x_i$. D'où

$$\overline{XY} = \sum_i \sum_j p'_i p''_{ij} x_i y_j = \sum_i p'_i x_i \left[\sum_j p''_{ij} y_j \right],$$

ce qui conduit à écrire

(56′′) $$\overline{XY} = \sum_i p'_i x_i (\mathfrak{M}_{X=x_i} Y),$$

en désignant par $\mathfrak{M}_{X=x_i} Y$ la valeur moyenne de Y *quand* $X = x_i$.

Cette expression de $\overline{XY}$ se simplifie évidemment dans un cas que pour cette raison, il y a lieu de considérer à part, celui où $\mathfrak{M}_{X=x_i} Y$ est indépendant de i. En représentant cette valeur commune par $\mathfrak{M}_X Y$,

(56^a) devient

$$\overline{XY} = \overline{X\mathfrak{M}_X Y} \tag{56b}$$

Or, on a

$$\mathfrak{M}_X Y = \sum_i p'_i \mathfrak{M}_X Y = \sum_i p'_i [\mathfrak{M}_{X=x_i} Y]$$

$$= \sum_i \sum_j p'_i p''_{ij} y_j = \sum_j \left[\sum_i p_{ij} \right] y_j = \sum_j p''_j y_j$$

D'où l'égalité, intéressante en soi

$$\mathfrak{M}_X Y = \overline{Y} \quad (^1), \tag{58}$$

et en vertu de (56^b),

$$\overline{XY} = \overline{X}\,\overline{Y} \quad (^2). \tag{56}$$

Le champ de validité de cette dernière formule se trouve donc étendu au cas où Y, sans être nécessairement indépendant de X, *est indépendant de* X « *en moyenne* », c'est-à-dire où la moyenne de Y dans la catégorie d'épreuves où X a une valeur donnée x_i garde la même valeur quelle que soit la valeur x_i choisie pour X.

Il en est, bien entendu, de même, quand X est indépendant « en moyenne » de Y.

Remarque. — On peut généraliser utilement une inégalité classique de Schwarz, en observant que, quel que soit le nombre certain λ, on a

$$\overline{X^2} + 2\lambda\overline{XY} + \lambda^2\overline{Y^2} = \mathfrak{M}(X + \lambda Y)^2 \geqq 0;$$

d'où

$$(\overline{XY})^2 \leqq \overline{X^2}\,\overline{Y^2} \tag{56c}$$

Valeur moyenne du carré d'une somme. — Au contraire, on n'est pas dans le domaine d'application de la formule (56) si l'on prend,

---

(1) *Voir* la note (2) p. 77, pour l'extension au cas correspondant à II.

(2) M. Cramér nous a fait observer que, en vertu de l'identité

$$\mathfrak{M}(X - \overline{X})(Y - \overline{Y}) = \overline{XY} - \overline{X}\,\overline{Y},$$

la condition nécessaire et suffisante pour qu'on ait l'égalité (56), c'est que le produit $(X - \overline{X})(Y - \overline{Y})$ ait une valeur moyenne.

Cette dernière propriété avait été antérieurement exprimée par nous en disant que X et Y sont « quadratiquement indépendants » (*voir* p. 73).

par exemple, $Y = X$. Il est facile de vérifier qu'on n'a pas, en général,

$$\overline{X^2} = \overline{X}^2.$$

Nous avons même vu plus haut que cette égalité ne peut avoir lieu que si X est presque toujours constant. Ceci résulte aussi de ce que, d'après les formules qu'on vient de démontrer,

$$(59) \qquad \overline{(X - \overline{X})^2} = \overline{X^2} - \overline{X}^2$$

En appelant ϖ un nombre certain arbitraire, cette formule peut aussi s'écrire

$$(59^{bis}) \qquad \boxed{\overline{(X - \varpi)^2} = \overline{(X - \overline{X})^2} + (\overline{X} - \varpi)^2,}$$

formule souvent utile, d'où l'on déduit que l'écart quadratique moyen de X avec un nombre certain ϖ atteint, lorsque ϖ varie, sa valeur minimum quand $\varpi = \overline{X}$ C'est pourquoi on pourra souvent appeler écart quadratique moyen de X, tout court, l'écart quadratique moyen de X et de sa valeur moyenne $\overline{X}$.

Soient $X_1, X_2, \ldots, X_n$, n variables aléatoires, $Y_k = X_k - \overline{X}_k$, l'accroissement de X_k à partir de sa valeur moyenne $\overline{X}_k$, et μ_k l'écart quadratique moyen de Y_k, μ' celui de $Y = Y_1 + \ldots + Y_n$. On a

$$0 \leqq \sum_{i,k} (Y_i - Y_k)^2 = n(Y_1^2 + \ldots + Y_n^2) - (Y_1 + \ldots + Y_n)^2.$$

La valeur moyenne du dernier membre est donc $\geqq 0$ et, par suite,

$$(60) \qquad \mu'^2 \leqq n(\mu_1^2 + \ldots + \mu_n^2)$$

Le signe d'égalité est d'ailleurs évidemment atteint quand les X_k sont identiques. Il est alors très intéressant de noter combien cette inégalité se trouve précisée dans le cas où les X_k sont indépendants. En effet, on a toujours

$$\mu'^2 = \overline{Y_1^2} + \ldots + \overline{Y_n^2} + 2\sum_{i,k} \overline{Y_i Y_k},$$

et, dans ce cas particulier,

$$\overline{Y_i . Y_k} = \overline{Y_i} . \overline{Y_k} = 0.$$

On obtient ainsi une formule qui semble due à Bienaymé (1, p. 166),

$$(61) \qquad \boxed{\mu'^2 = \mu_1^2 + \ldots + \mu_n^2.}$$

Ainsi, le carré de l'écart quadratique moyen d'une somme de variables aléatoires *indépendantes* X_i est égal à la somme des carrés des écarts quadratiques moyens de ces variables respectives X_i. C'est là une des propriétés les plus utiles de l'écart quadratique moyen.

En particulier, si nous désignons par X_k la valeur aléatoire prise à une $k^{\text{ième}}$ épreuve par une même variable aleatoire X, et si nous supposons indépendants les résultats des épreuves, on pourra prendre pour μ_k l'écart quadratique moyen μ de X, et l'on aura $\mu' = \mu\sqrt{n}$. Si l'on appelle λ l'écart quadratique moyen de la moyenne arithmétique $v_n = \frac{X_1 + \ldots + X_n}{n}$, on aura $\overline{v_n} = \overline{X}$ et $\lambda = \frac{\mu'}{n}$; d'où $\lambda = \frac{\mu}{\sqrt{n}}$. Ainsi, l'écart quadratique moyen de v_n varie comme $\frac{1}{\sqrt{n}}$ quand n varie.

Nous allons établir une généralisation à l'ordre quatre qui nous sera utile plus tard.

On a

$$(Y_1 + \ldots + Y_n)^4 = \sum_i Y_i^4 + 6\sum_{i,k} Y_i^2 Y_k^2 + 24 \sum_{i,j,k,l} Y_i Y_j Y_k Y_l$$
$$+ 4\sum_{i,k} Y_i^3 Y_k + 12 \sum_{i,j,k} Y_i^2 Y_j Y_k$$

Quand les X_k sont indépendants, les moyennes des termes où figure un Y_k au premier degré sont nulles. On a donc, dans ce cas,

$$\overline{(Y_1 + \ldots + Y_n)^4} = \sum_i \overline{Y_i^4} + 6\sum_{i,k} \overline{Y_i^2}.\overline{Y_k^2}$$

Dans le cas particulier où les X_k sont les valeurs, au cours d'épreuves indépendantes, d'une variable indépendante X dont les écarts moyens d'ordres 2 et 4 avec $\overline{X}$ sont μ et ν, l'écart moyen ν' d'ordre 4 de la somme $X_1 + \ldots X_n$ sera donné par

$$\nu'^4 = n\nu^4 + 3n(n-1)\mu^4 \tag{62}$$

Comme on a vu que $\mu \leqq \nu$, on a donc

$$\nu' \leqq \nu \sqrt[4]{n(3n-2)}.$$

L'écart moyen $\rho = \frac{\nu'}{n}$ d'ordre 4, de $(\overline{v_n} - \overline{X})$ sera donc tel que

$$\rho \leqq \nu \sqrt[4]{\frac{3n-2}{n^3}}. \tag{63}$$

Généralisation de l'égalité de Bienaymé. — L'importante formule (61) a été établie dans l'hypothèse où les X_j sont indépendants. Mais elle est plus générale. Pour qu'elle soit valable, il suffit que les $\overline{Y_i Y_k}$ $(i \neq k)$, soient tous nuls.

Nous dirons que des variables aléatoires X_i sont *quadratiquement indépendantes* si les moyennes $\mathfrak{M}\{(X_i - \overline{X_i})(X_k - \overline{X_k})\}$ sont toutes nulles $(i \neq k)$. On peut rattacher cette notion à une autre plus connue.

On a, en vertu de (56^c)

$$0 \leqq (\overline{Y_i Y_k})^2 \leqq \mu_i^2 \mu_k^2,$$

par conséquent : ou bien $\mu_i \mu_k = 0$, et alors $\overline{Y_i Y_k} = 0$ ou bien $\mu_i \mu_k \neq 0$ et alors on peut écrire

$$\overline{Y_i Y_k} = r_{ik} \mu_i \mu_k,$$

en posant

$$r_{ik} = \frac{\mathfrak{M}\{(X_i - \overline{X_i})(X_k - \overline{X_k})\}}{\sqrt{[\mathfrak{M}(X_i - \overline{X_i})^2][\mathfrak{M}(X_k - \overline{X_k})^2]}}$$

et l'on aura

$$0 \leqq |r_{ik}| \leqq 1$$

Dans l'expression de r_{ik} on reconnaît la quantité appelée jusqu'ici (mais improprement) « coefficient de corrélation » de X_i et de X_k.

Observons que si $\mu_i = 0$, on a $\mathfrak{M}(X_i - \overline{X_i})^2 = 0$, c'est-à-dire que X_i est égal à un nombre certain $\overline{X_i}$ ou bien est « presque sûrement » égal à celui-ci.

Ainsi, pour que des variables aléatoires X_i vérifient l'égalité (61) de Bienaymé, il suffit qu'elles soient quadratiquement indépendantes. c'est-à-dire qu'en laissant éventuellement de côté celles des X_i qui seraient certaines (ou presque sûrement certaines) les coefficients dits de corrélation des autres variables prises 2 à 2 soient tous nuls.

Il est facile de voir que cette condition est effectivement plus générale que celle de l'indépendance. C'est dire que l'utilité de la considération d'un coefficient dit de corrélation est limitée précisément, dans l'exemple actuel, au cas où celui-ci ne mesure pas bien le degré de corrélation (ou de dépendance) et ne justifie pas son nom. Nous avons insisté ailleurs (Fréchet, 159) sur ce point et proposé pour ce coefficient le nom de « coefficient de linéarité », parce qu'il n'atteint, en valeur absolue, sa plus grande valeur, l'unité, que s'il

existe entre X_i et X_k une relation linéaire (presque sûrement) (c'est là une propriété bien connue et dont on s'assure facilement).

Un cas particulier important où l'égalité (61) subsiste a été signalé par M. Cantelli (8). On peut le caractériser en disant que c'est le cas où les variables X_i sont indépendantes « en moyenne », c'est-à-dire où pour tout couple X_i, X_k, l'une au moins de ces deux variables est indépendante « en moyenne » (p. 70) de l'autre. On a, en effet, vu que, dans ce cas,

$$\overline{Y_i Y_k} = \overline{Y_i}\,\overline{Y_k} = 0$$

Fonctions des probabilités totales de la somme et du produit de deux variables aléatoires indépendantes. — Connaissant les fonctions des probabilités totales $F(x)$, $G(y)$ de deux variables aléatoires X, Y, peut-on en déduire l'expression de la fonction des probabilités totales d'une fonction déterminée (par exemple, de la somme) de X et de Y ? Il est facile de reconnaître que la réponse est négative si les variables X, Y ne sont pas indépendantes, quand on ne connaît pas la dépendance qui existe entre ces deux variables. Soit, par exemple, à trouver la fonction des probabilités totales de $Z = X + Y$. Si X, Y sont liées par la relation $X + Y = 0$, on aura $Z = 0$; si X et Y sont liées par la relation $Y - X = 0$, on aura $Z = 2X$. Il est clair que, X restant le même, Y, qui ne sera pas le même d'un cas à l'autre, pourra garder, comme X, la même fonction des probabilités totales. (Par exemple, c'est ce qui aura lieu, si X suit une loi symétrique comme la loi de Laplace et si, en outre, $\overline{X} = 0$.) Pourtant, la fonction des probabilités totales de Z sera différente d'un cas à l'autre [1].

Au contraire, la réponse au problème posé ci-dessus est affirmative dans le cas où X et Y sont deux variables indépendantes.

I. *Somme.* — Considérons d'abord le cas de la somme $Z = X + Y$ et cherchons sa fonction des probabilités totales

$$S(z) = \{\text{Pr.}[X + Y < z]\}.$$

On peut considérer $S(z)$ comme représentant la probabilité pour

---

[1] Dans ces deux cas, Y est déterminé certainement quand on donne X. Il peut aussi arriver que la fonction des probabilités totales de Y ne soit pas la même quand X est donné ou quand X est non donné, sans que la connaissance de X détermine Y.

que le point aléatoire de coordonnées X, Y soit situé au-dessous de la droite $x+y=z$. Soit $\ldots, x_{-m}, x_{-m+1}, \ldots, x_{-1}, x_0, x_1, \ldots, x_n, \ldots$, une suite de nombres croissant de $-\infty$ à $+\infty$ et dont les intervalles sont tous inférieurs à un même nombre positif δ. La région $x+y<z$ comprend une certaine région r et est comprise dans une certaine région R définies ainsi qu'il suit.

La région r est l'ensemble des régions r_j définies par

$$x_{j-1} \leqq x < x_j, \qquad y < z - x_j$$

La région R est l'ensemble des régions R_k définies par

$$x_{k-1} \leqq x < x_k, \qquad y < z - x_{k-1}$$

Puisque X et Y sont indépendantes, la probabilité pour que le point (X, Y) soit sur r_j est

$$[F(x_j) - F(x_{j-1})]\, G(z - x_j)$$

et la probabilité pour que (X, Y) soit sur R_k est

$$[F(x_k) - F(x_{k-1})]\, G(z - x_{k-1})$$

Puisque les régions r_j sont sans point commun, de même que les régions R_k, on peut leur appliquer le théorème des probabilités totales. Finalement, on a

$$t \leqq S(z) \leqq T,$$

en posant

$$t = \sum_{j=-\infty}^{j=+\infty} G(z - x_j) \quad [F(x_j) - F(x_{j-1})],$$

$$T = \sum_{j=-\infty}^{j=+\infty} G(z - x_{j-1})[F(x_j) - F(x_{j-1})],$$

et ces deux séries extrêmes qui sont à termes $\geqq 0$ sont convergentes; et, même, leurs sommes sont $\leqq 1$.

L'intégrale de Stieltjes $\int_{-\infty}^{+\infty} G(z-x)\, dF(x)$ peut se définir (p. 39) comme la limite commune de ces deux quantités t et T, quand on fait tendre δ vers zéro. On a donc

$$(64) \qquad \boxed{S(z) = \int_{-\infty}^{+\infty} G(z-x)\, dF(x)}$$

et de même

$$(65) \qquad S(z) = \int_{-\infty}^{+\infty} F(z-y)\, d\,G(y)$$

Autrement dit, $S(z)$ *est la valeur moyenne* pour z donné, *de la variable aléatoire* $F(z-Y)$ ou *de la variable aléatoire* $G(z-X)$.

Le raisonnement précédent est fondé sur l'existence d'une limite commune quand $\delta \to 0$ de deux sommes (absolument convergentes)

$$\sum_{j=-\infty}^{j=+\infty} g(x_j)\,[F(x_j)-F(x_{j-1})], \quad \sum_{j=-\infty}^{j=+\infty} g(x_{j-1})\,[F(x_j)-F(x_{j-1})]$$

où $F(x)$ et $g(x)$ sont deux fonctions monotones bornées A vrai dire, ce point mérite qu'on s'y arrête un instant.

Il n'y a pas de difficulté quand la fonction $g(x)$ est supposée continue Étant monotone et bornée, elle est alors (p 34) uniformément continue sur l'ensemble des points de la droite illimitée. A tout $\varepsilon > 0$ correspond un nombre ω tel que $|g(x')-g(x'')| < \varepsilon$ pour $|x'-x''| < \omega$ Alors en prenant $\delta < \omega$, la différence des sommes ci-dessus sera inférieure à

$$\varepsilon \sum_{j=-\infty}^{j=+\infty} |F(x_j)-F(x_{j-1})| = \varepsilon$$

Il reste à justifier l'hypothèse quand $g(x)$ est discontinue Dans le cas où $g(x) = G(z-x)$, on observera que la fonction $F(x)$ est continue à gauche et que la fonction $g(x)$ est continue à droite Or on peut démontrer (Fréchet, 183) que

Si $g(x)$ et $F(x)$ sont deux fonctions monotones bornées, qui en aucun point ne sont discontinues du même côté, les deux sommes ci-dessus tendent vers la même limite quand $\delta \to 0$ pourvu que les modes de division employés satisfassent à la condition suivante : que tout point de discontinuité commun à g et à F finisse par appartenir à tous les modes de division considérés à partir d'une valeur convenable de δ. On peut appliquer cette condition au cas de la fonction $g(x) = G(z-x)$ rien n'empêche, pour une valeur déterminée (arbitraire) de z, d'assujettir la suite des x_i à cette condition

Par suite la formule (64) se trouve établie dans tous les cas [1]. Et de même pour (65).

Une démonstration differente de la formule (64) a été donnée par M. Paul Lévy (1, p. 188). Cette démonstration est basée sur le lemme suivant : si E_X est un événement dépendant d'une variable

[1] M Glivenko (2) est arrivé au même résultat par une méthode différente.

aléatoire X et si $\lambda(x)$ est la probabilité pour que E_X ait lieu quand X a une valeur déterminée x, alors la probabilité λ pour que E_X ait lieu, sans qu'on se donne d'avance la valeur de X, est égale à la valeur moyenne de $\lambda(X)$. Un tel lemme est assez naturel en soi : quand X ne peut prendre qu'un nombre fini de valeurs il devient facile à démontrer [1]. Mais dans le cas contraire, il n'en n'est plus ainsi [2]. Si l'on admet ce lemme au moins quand $\lambda(x)$ est une fonction monotone de x, alors il suffit d'observer que la probabilité pour que $x+Y<z$ est égale à $G(z-x)$, fonction monotone de x. Par suite, la probabilité $H(z)$ pour que $X+Y<z$ sera bien la moyenne de $G(z-X)$.

II. *Produit.* — On opère de façon analogue pour obtenir la fonction des probabilités totales $P(u)$ du produit $U=XY$. Considérons d'abord le cas où $u>0$. La région où $XY<u$ est formée par l'inté-

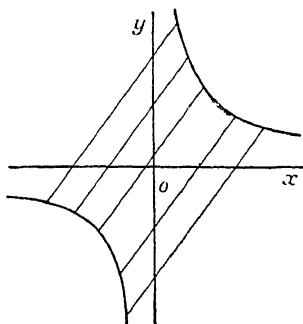



Fig. 2

rieur (représenté sur la figure 2 par des hachures) de l'hyperbole $xy=u$. Elle est aussi comprise entre deux régions r et R, sommes respectives de régions r_j et R_k. Ici, nous supposerons $x_0=0$; alors si $x_{j-1}\geqq 0$, r_j est définie par

$$x_{j-1}\leqq x<x_j,\qquad y<\frac{u_j}{x_j},$$

(1) Il suffit d'appliquer la formule (58), p. 70, au cas où Y ne prend que les valeurs 0 ou 1.

(2) M. Paul Levy a bien voulu reconnaître l'exactitude de l'observation que je lui avais communiquée à ce sujet et m'annoncer qu'il publie (7) une démonstration complète de ce lemme.

et R_j par

$$x_{j-1} \leqq x < x_j, \qquad y < \frac{u}{x_{j-1}}$$

et si $x_j \leqq 0$, r_j est défini par

$$x_{j-1} \leqq x < x_j, \qquad y > \frac{u}{x_{j-1}},$$

et R_j par

$$x_{j-1} \leqq x < x_j, \qquad y > \frac{u}{x_j}.$$

On a, en prenant pour les $x_j \neq 0$ des points x où $G\left(\frac{u}{x}\right)$ est continue

$$\sum_{x_j=-\infty}^{x_j=0} \left[1 - G\left(\frac{u}{x_{j-1}}\right)\right] [F(x_j) - F(x_{j-1})] + \sum_{x_j=x_1}^{x_j=+\infty} G\left(\frac{u}{x_j}\right) [F(x_j) - F(x_{j-1})]$$

$$\leqq P(u) \leqq \sum_{x_j=+\infty}^{x_j=0} \left[1 - G\left(\frac{u}{x_j}\right)\right] [F(x_j) - F(x_{j-1})]$$

$$+ \sum_{x_j=x_1}^{x_j=-\infty} G\left(\frac{u}{x_{j-1}}\right) [F(x_j) - F(x_{j-1})],$$

où l'on remplace $G\left(\frac{u}{x_0}\right)$ dans le dernier membre par o dans le premier Σ, — où les x sont $\leqq 0$ et par 1 dans le second Σ, — où les x sont $\geqq 0$.

On en déduit, comme plus haut, que si $u > 0$,

$$(66) \qquad P(u) = \int_{-\infty}^{0} \left[1 - G\left(\frac{u}{x}\right)\right] dF(x) + \int_{0}^{+\infty} G\left(\frac{u}{x}\right) dF(x).$$

Pour $u < 0$, la formule est encore valable; le raisonnement est analogue mais s'applique à la région $XY < u\,(< 0)$ représentée par des hachures sur une figure symétrique à la figure 2 par rapport à Oy. Enfin, pour $u = 0$, on voit directement que

$$P(0) = [1 - G(0)]\, F(0) + G(0)\,[1 - F(0)],$$

si $F(x)$ et $G(y)$ sont continus pour $x = 0$ et $y = 0$, de sorte que la formule (66) est dans ce cas applicable quel que soit u.

On peut l'interpréter en disant que $P(u)$ est la valeur moyenne d'une variable aléatoire égale à $1 - G\left(\frac{u}{X}\right)$ quand $X < 0$, à $G\left(\frac{u}{X}\right)$ quand $X > 0$ et égale quand $X = 0$, soit à o si $u > 0$, soit à 1 si $u < 0$.

Cas des variables dépendantes. — S'il n'est plus possible dans ce cas de déterminer $S(z)$ connaissant seulement $F(x)$ et $G(y)$, on peut cependant en reprenant la démonstration faite dans le cas des variables indépendantes aboutir à une formule utile quoique moins simple.

En employant les notations précédentes, on voit que la probabilité pour que le point (X, Y) soit sur r_j est

$$[F(x_j) - F(x_{j-1})] g(z - x_j, x_{j-1}, x_j)$$

en appelant $g(y, x', x'')$ la probabilité pour que $Y < y$ *quand* $x' \leqq X < x''$. De même la probabilité pour que (X, Y) soit sur R_k est

$$[F(x_k) - F(x_{k-1})] g(z - x_{k-1}, x_{k-1}, x_k)$$

D'où

$$s = \sum_{j=-\infty}^{j=+\infty} [F(x_j) - F(x_{j-1})] g(z - x_j, \quad x_{j-1}, x_j)$$

$$\leqq S(z) \leqq \sum_{j=-\infty}^{j=+\infty} [F(x_j) - F(x_{j-1})] g(z - x_{j-1}, x_{j-1}, x_j) = s'$$

On a

$$0 \leqq S(z) - s \leqq s' - s$$

$$= \sum_{j=-\infty}^{j=+\infty} [F(x_j) - F(x_{j-1})][g(z - x_{j-1}, x_{j-1}, x_j) - g(z - x_j, x_{j-1}, x_j)]$$

Soit ω la borne supérieure de l'oscillation de $g(y, x', x'')$ quand, x' et x'' restant fixes, y varie dans un intervalle de longueur $< \delta$. Quand x', x'' varient, ω variera.

Soit $\Omega(\delta)$ la borne supérieure de ω quand x'. x'' varient, mais de sorte que $x'' - x' < \delta$.

Observons, en passant, avec M. Cramér, qu'on a pour tout $h > 0$

$$S(z + h) - S(z) \leqq \sum_j [F(x_j) - F(x_{j-1})]$$
$$\times [g(z + h - x_{j-1}, x_{j-1}, x_j) - g(z - x_j, x_{j-1}, x_j)],$$

d'où

$$0 \leqq S(z + h) - S(z) \leqq \sum_j [F(x_j) - F(x_{j-1})] \Omega(h + \varepsilon) \leqq \Omega(\delta) \qquad \text{pour } 0 < h < \delta,$$

en prenant les $x_j - x_{j-1} < \varepsilon < \delta - h$ et que par suite, le module de

continuité $\theta(\delta)$ de $S(z)$, c'est-à-dire le maximum de $S(z')-S(z)$, quand $0 > z'-z < \delta$, est au plus égal à $\Omega(\delta)$.

On aura d'autre part

$$0 \leqq S(z) - s \leqq \Omega(\delta).$$

Si donc, $\Omega(\delta)$ tend vers zéro avec δ, s devra tendre vers $S(z)$. Pour définir un cas général où il en est ainsi et en même temps simplifier l'expression de la limite de s, il est naturel d'introduire la probabilité $G_x(y)$ pour que $Y < y$ *quand* $X = x$ et d'écrire

$$\begin{aligned} & g(z-x_{j-1}, x_{j-1}, x_j) - g(z-x_j, x_{j-1}, x_j) \\ = & \left[g(z-x_{j-1}, x_{j-1}, x_j) - G_{x_{j-1}}(z-x_{j-1})\right] \\ & - \left[g(z-x_j, \quad x_{j-1}, x_j) - G_{x_{j-1}}(z-x_j)\right] \\ & + \left[G_{x_{j-1}}(z-x_{j-1}) - G_{x_{j-1}}(z-x_j)\right] \end{aligned}$$

Supposons maintenant :

1° que $G_x(y)$ soit une fonction de y continue et même que les fonctions de y, $G_x(y)$ obtenues en faisant varier le paramètre x, soient « également » continues, c'est-à-dire que l'oscillation maximum $\varepsilon(x)$ de $G_x(y)$ quand y varie dans un intervalle de valeurs de y de longueur $< \delta$, ait quand x varie une borne supérieure $\varepsilon(\delta)$ tendant vers zéro avec δ;

2° que $g(y, x', x'')$ tende vers $G_{x'}(y)$ uniformément quand x'' tend vers x' ou, plus précisément, qu'en désignant par $\eta(\delta)$ le maximum de $|g(y, x', x'') - G_{x'}(y)|$ lorsque y varie arbitrairement et que x', x'' varient arbitrairement, mais de sorte que $|x''-x'| < \delta$, alors $\eta(\delta)$ tende vers zéro avec δ. Dans ces conditions, on a

$$\Omega \leqq 2\,\eta(\delta) + \varepsilon(\delta),$$

d'où $\lim\limits_{\rho \to 0} \Omega = 0$ et par suite $S(z) = \lim\limits_{\rho \to 0} s$.

Or, soit

$$s_1 = \sum_{j=-\infty}^{j=+\infty} [F(x_j) - F(x_{j-1})]\, G_{x_{j-1}}(z-x_j),$$

c'est une série convergente à termes $\geqq 0$ et l'on a

$$|s_1 - s| \leqq \sum_{j=-\infty}^{j=+\infty} [F(x_j) - F(x_{j-1})]\,|G_{x_{j-1}}(z-x_j) - g(z-x_j, x_{j-1}, x_j)| \leqq \eta(\delta),$$

de sorte que $s_1 - s$ tend vers zéro avec δ et, par suite, $S(z) = \lim_{\delta \to 0} s_1$.

Finalement, on voit qu'on a

$$S(z) = \int_{-\infty}^{+\infty} G_1(z-x)\, dF(x). \tag{64 bis}$$

M. Paul Lévy (4, p. 94) a bien voulu me communiquer, au sujet de cette formule, une intéressante remarque généralisant un résultat qu'il avait publié antérieurement (1, p. 188-189).

On déduit, en effet, de (64 *bis*)

$$S(z') - S(z) = \int_{-\infty}^{+\infty} [G_1(z'-x) - G_1(z-x)]\, dF(x),$$

d'où, si $0 \leqq z' - z \leqq \delta$,

$$0 \leqq S(z') - S(z) \leqq \int_{-\infty}^{+\infty} \varepsilon(\delta)\, dF(x) = \varepsilon(\delta)$$

Si donc on désigne encore par $\theta(\delta)$ le module de continuité de $S(z)$, on aura $\theta(\delta) \leqq \varepsilon(\delta)$.

De sorte que la fonction des probabilités totales de la somme de X et de Y est au moins aussi continue ou plus continue que la fonction $G_1(y)$ de y. Un énoncé antérieur de M. Paul Lévy était d'une portée moins générale, car il ne s'appliquait qu'au cas de variables indépendantes, mais il était plus simple car dans ce cas $G_1(y)$ est indépendant de x : la fonction des probabilités totales de la somme de deux variables aléatoires indépendantes X, Y est toujours continue quand la fonction des probabilités totales de l'une de ces variables aléatoires est continue et, dans ce cas, la première est au moins aussi continue et en général plus continue que la seconde.

Observons aussi que si l est la longueur d'une suite dénombrable d'intervalles quelconques (ζ_r, ζ'_r) ne chevauchant pas, on a

$$\sum_r [S(\zeta'_r) - S(\zeta_r)] = \int_{-\infty}^{+\infty} \left\{ \sum_r [G_1(\zeta'_r - x) - G_1(\zeta_r - x)] \right\} dF(x)$$

Si maintenant $G_1(z-x)$ est absolument continue (¹) en z, alors

---

(¹) Une fonction $\Phi(x)$ est dite (suivant Vitali) absolument continue si la somme $\sum_r |\Phi(x'_r) - \Phi(x_r)|$ prise sur un ensemble dénombrable d'intervalles deux

quand l tend vers zéro, l'accolade tend vers zéro et cela uniformément quand x varie. Par suite, il en est de même du premier membre. $S(z)$ est aussi absolument continue.

En particulier (P. Lévy, 1, p. 189) si la fonction des probabilités totales de l'une des variables aléatoires indépendantes X, Y est absolument continue (définition p. 33), il en est de même de la fonction des probabilités totales de leur somme $X + Y$.

Il en résulte que si la fonction des probabilités totales de Y, par exemple, est de la forme

$$G(y) = \int_{-\infty}^{y} g(t)\, dt,$$

celle de $X + Y$ sera de la forme

$$S(z) = \int_{-\infty}^{z} s(t)\, dt$$

avec

$$s(z) = \int_{-\infty}^{+\infty} g(z - x)\, dF(x)$$

Points aléatoires. Points moyens. — Les applications mêmes du Calcul des Probabilités imposent une généralisation des notions de nombre aléatoire et de valeur moyenne. Il est rare qu'un événement aléatoire soit caractérisé par un seul paramètre. S'il est caractérisé par deux paramètres, on pourra le représenter par un point du plan. Si, plus généralement, il est caractérisé par n paramètres, on pourra employer un langage figuré et le considérer comme définissant un point de l'espace à n dimensions. Mais on peut avoir affaire à des événements plus complexes encore. On peut avoir, par exemple, à considérer des courbes tracées au hasard. Dans ce cas on pourra considérer l'élément aléatoire à étudier comme un point d'un espace plus général à une infinité de dimensions.

On est ainsi conduit à la conception d'un point aléatoire appartenant à un espace dit abstrait en ce sens qu'aucune limitation n'est attachée à la nature de ses éléments.

L'utilité de cette conception va se révéler quand il s'agira de géné-

---

à deux disjoints (x_r, x_r') tend vers zéro quand la somme des longueurs de ces intervalles tend vers zéro. C'est la condition nécessaire et suffisante pour que $\Phi(x)$ soit une intégrale de M. Lebesgue $\int \varphi(x)\, dx$.

raliser la notion de valeur moyenne. Si, par exemple, on considere un nombre X, bien déterminé quand on connaît la position d'un point aléatoire x, X sera une fonction numérique bien déterminée de x

$$X = \varphi(x).$$

Supposons, d'autre part, qu'il existe une probabilité déterminée $\varpi(v)$ pour que x appartienne à un sous-ensemble déterminé v de l'ensemble V des positions possibles de x dans l'espace abstrait considéré. Alors, pour des motifs à peu près évidents et que nous développerons du reste plus tard, on pourra calculer la valeur moyenne de X sur V au moyen de l'intégrale généralisée

$$\bar{X} = \int_V \varphi(x)\, d\varpi\,(v)$$

Mais il faut préciser ce que représente exactement l'intégrale du second membre. Quand V appartient à l'espace à n dimensions, on pourra prendre l'intégrale au sens de Radon (1). Mais nous avons montré ailleurs (55) qu'on peut étendre cette définition au cas le plus général, celui où V appartient à un espace abstrait quelconque sans qu'il soit même nécessaire d'y supposer donnée une définition de la limite d'une suite de ses éléments. Nous nous contenterons provisoirement ici de ces indications en renvoyant pour la définition précise de l'intégrale sur un ensemble abstrait à notre Mémoire et, pour les applications, au livre de M. Kolmogoroff (4, p. 1, 33) aux yeux duquel cette conception de l'intégrale sur un ensemble abstrait est indispensable en Calcul des Probabilités. Il y aura d'ailleurs lieu aussi de généraliser (Gateaux, J.-L. Destouches) la notion de point moyen qui jouera dans l'étude d'un point aléatoire abstrait le rôle du centre de gravité pour les points matériels.

SECTION III. ÉPREUVES RÉPÉTÉES.

Valeurs moyennes des fréquences. — Soit $f = \frac{r}{n}$ la fréquence d'un événement fortuit E au cours de n épreuves où il s'est répété r fois. Si E a une probabilité constante p, la valeur moyenne de f [1] pourrait

---

[1] Bien entendu, si E est défini sur une catégorie C d'épreuves, la valeur

se calculer par la formule

$$\bar{f} = \frac{0}{n} q^n + \frac{1}{n} C_n^1 q^{n-1} p + \ldots + \frac{l}{n} C_n^l q^{n-l} p^l + \ldots + \frac{n}{n} p^n$$

Mais les théoremes établis plus haut fournissent une expression plus simple. On peut écrire $r = X_1 + \ldots + X_n$, où X_k est égal à 1 ou 0 suivant qu'à la $k^{\text{ième}}$ épreuve E se produit ou non. On a évidemment par définition de la valeur moyenne $\overline{X_k} = 1 \times p + 0 \times q = p$. D'où, puisque $\bar{f} = \frac{\bar{r}}{n}$

$$\boxed{\bar{r} = np, \qquad \bar{f} = p} \tag{67}$$

La même méthode fournit facilement d'autres expressions. Soient μ' et μ les écarts quadratiques moyens de r et de f. On a évidemment $\mu = \frac{\mu'}{n}$ et l'écart quadratique moyen μ_1 de X_1, est donné, par définition, par

$$\mu_1^2 = (1 - \overline{X_1})^2 p + (0 - \overline{X_1})^2 q = pq$$

Donc, en vertu de l'égalité de Bienaymé (61),

$$\mu'^2 = npq,$$

d'où

$$\mu' = \sqrt{npq}, \qquad \mu = \sqrt{\frac{pq}{n}}.$$

Calculons aussi les moments d'ordre 3 et 4 de la déviation $f - p$. On a, en posant $Y_k = X_k - \overline{X_k} = X_k - p$.

$$\begin{aligned}\overline{(r - np)^3} &= \overline{(Y_1 + \ldots + Y_n)^3} = (\overline{Y_1^3} + \ldots + \overline{Y_n^3}) + 3\sum_{l,k} \overline{Y_l^2}\,\overline{Y_k} + 6\sum_{l,j,k} \overline{Y_l}\,\overline{Y_j}\,\overline{Y_k} \\ &= n\overline{Y_1^3} = n\{p(1-p)^3 + q(0-p)^3\} \\ &= npq(q-p).\end{aligned}$$

D'où

$$\overline{(f - p)^3} = \frac{pq(q-p)}{n^2}. \tag{68}$$

D'autre part, les écarts moyens ν' et ν d'ordre 4 de r et f, tels

---

moyenne de f est calculée dans la catégorie complexe C_n formée des groupes de n épreuves.

que $\nu = \frac{\nu'}{n}$, s'obtiennent en calculant

$$\overline{Y_1^4} = p(1-p)^4 + q(0-p)^4 = pq(p^2+q^2-pq)$$

et en utilisant la formule (62) avec changement de notation. On en tire

$$\nu'^4 = npq(p^2+q^2-pq) + 3n(n-1)p^2q^2.$$

D'où

$$(69)\qquad \nu = \sqrt[4]{\frac{pq}{n^3}[p^2+q^2+(3n-4)pq]} = \sqrt[4]{\frac{pq}{n^3}[1+3(n-2)pq]}$$

Nouveau calcul de l'écart moyen. — Chemin faisant, calculons aussi l'écart moyen σ' de f avec p, $\sigma' = \overline{|f-p|}$. Ce calcul a été fait par Bertrand (1, p. 82). On peut même étendre (Fréchet, 176) la méthode de Bertrand de façon à obtenir l'écart moyen σ_w de f à partir d'une valeur certaine quelconque w (et non à partir de p).

Si w est $\leqq 0$ ou est $\geqq 1$, on a

$$\sigma_w = \mathfrak{M}|f-w| = |\mathfrak{M}(f-w)| = |p-w|,$$

puisque $f-w$ garde un signe constant.

Si $0 < w < 1$, comme on a

$$\sigma_w = \sum_{r=0}^{r=n} \left| \frac{r}{n} - w \right| T_r$$

avec

$$T_r = C_n^r p^r q^{n-r},$$

on est conduit pour déterminer le signe de $\frac{r}{n} - w$ à désigner par ρ l'entier tel que

$$\frac{\rho}{n} \leqq w < \frac{\rho+1}{n}.$$

D'où

$$n\sigma_w = \sum_{r \leqq \rho} (nw - r)T_r + \sum_{r>\rho} (r - nw)T_r$$

Pour introduire l'artifice de Bertrand, on écrira

$$n\sigma_w = B + nA$$

avec

$$B = \sum_{r \leqq \rho} (np - r) T_r + \sum_{r > \rho} (r - np) T_r,$$

$$A = \sum_{r \leqq \rho} (w - p)\, T_r + \sum_{r > \rho} (p - w)\, T_r$$

En posant

$$P(\rho) = \sum_{r \leqq \rho} T_r,$$

c'est-à-dire en désignant par $P(\rho)$ la probabilité que l'événement considéré se répète ρ fois au plus (dans les n épreuves), on aura

$$A = (w - p)[2P(\rho) - 1].$$

Pour calculer B, on observe, avec Bertrand, que si l'on considère, pour un instant, p et q comme des variables indépendantes, on a

$$\frac{dT_r}{dq} - \frac{dT_r}{dp} = \frac{np - r}{pq} T_r;$$

d'où

$$B = pq \left\{ \left[\sum_{r \leqq \rho} \left(\frac{dT_r}{dq} - \frac{dT_r}{dp} \right) \right] + \left[\sum_{r > \rho} \left(\frac{dT_r}{dp} - \frac{dT_r}{dq} \right) \right] \right\}.$$

Or les termes de chaque crochet se réduisent deux par deux à l'exception d'un seul. Le premier crochet, par exemple, est

$$\begin{aligned} &nq^{n-1} + C_n^1 (n-1) q^{n-2} p + \ldots + C_n^{\rho-1} (n - \rho + 1) q^{n-\rho} p^{\rho-1} \\ &\quad + C_n^{\rho} (n - \rho) q^{n-\rho-1} p^{\rho} - C_n^1 q^{n-1} - C_n^2\, 2pq^{n-2} - \ldots - C_n^{\rho}\, \rho p^{\rho-1} q^{n-\rho} \\ &= C_n^{\rho} (n - \rho) q^{n-\rho-1} p^{\rho} = \frac{n - \rho}{q} T_\rho \end{aligned}$$

De même, le second crochet se réduit à

$$C_n^{\rho+1} (\rho + 1) p^{\rho} q^{n-\rho-1} = \frac{n - \rho}{q} T_\rho,$$

d'où

$$B = 2p(n - \rho) T_\rho.$$

Il en résulte l'*expression cherchée*

$$\sigma_w = 2 \left\{ p \left(1 - \frac{\rho}{n} \right) T_\rho + (w - p) \left[P(\rho) - \frac{1}{2} \right] \right\} \tag{70}$$

pour

$$\frac{\rho}{n} \leqq \varpi < \frac{\rho + 1}{n}.$$

La fonction σ_ϖ étant continue, cette expression doit coïncider (et c'est ce qu'on vérifie) pour $\varpi = 0$ et 1 avec l'expression $\sigma_\varpi = |\varpi - p|$ obtenue pour $\varpi \leqq 0$ ou $\varpi \geqq 1$.

En particulier, on obtient l'écart de f avec p

$$(71) \qquad \sigma' = \sigma_p = 2p\left(1 - \frac{R'}{n}\right) T_{R'},$$

où R' est la partie entière de np. En posant $np = R' + \theta'$, on peut écrire aussi

$$\sigma' = 2p\left(q + \frac{\theta'}{n}\right) T_{R'}$$

A la vérité, telle n'est pas précisément l'expression donnée par Bertrand comme résultat de ses calculs, eux-mêmes rigoureusement exacts. Il prend pour valeur de σ' l'expression asymptotiquement équivalente

$$(72) \qquad \sigma' \simeq 2pq T_R,$$

où R est la répétition la plus probable

Or $q + \frac{\theta'}{n}$ est supérieur à q, sauf dans le cas exceptionnel où np est un nombre entier et où par suite $q + \frac{\theta'}{n} = q$. D'autre part, comme nous l'avons fait observer ailleurs (F. et H., p. 196), la répétition la plus probable, soit R, est égale à la partie entière de $(n+1)p$ (et non pas nécessairement, comme on le voit parfois affirmer, à la partie entière R' de np). On peut seulement dire que R est égal à R' ou à $R' + 1$. (Par exemple, si $n = 11$ et $p = 0,8$, on a

$$np = 8,8, \qquad (n+1)p = 9,6;$$

d'où

$$R' = 8, \qquad R = 9 = R' + 1.)$$

Donc $T_{R'}$, est, soit égal à T_R, soit inférieur à T_R.

On voit ainsi que les deux erreurs commises en substituant $2pqT_R$ à l'expression exacte de σ' se compensent en partie. Néanmoins il peut arriver que l'erreur finale ne soit pas négligeable. Par exemple,

pour $n=7$, $p=0,48$, le rapport de la valeur approchée à la valeur exacte est de $\frac{28}{31}$, soit une erreur relative d'environ $\frac{1}{10}$.

Il faut d'ailleurs observer qu'une troisième cause d'erreur se présente du fait que σ' n'est pas ce qu'il conviendrait d'appeler écart moyen, tout court, de f. Il serait naturel d'appeler ainsi le minimum de σ_w quand w varie. Or on sait que celui-ci a lieu, non quand $w=p$, mais quand w est égal à la médiane m ou (l'une des médianes) de f.

On vérifie directement sur l'expression (70) de σ_w, que la valeur minimum $\sigma=\sigma_m$ de σ_w a lieu quand nw prend la valeur M de ρ telle que

$$P(M-1)<\frac{1}{2}, \qquad P(M)\geqq\frac{1}{2}, \qquad \text{avec} \quad m=\frac{M}{n}.$$

On a, alors

$$\sigma=\sigma_m=2\left\{p(1-m)T_M+(m-p)\left[P(M)-\frac{1}{2}\right]\right\}.$$

On peut observer qu'on a

$$0\leqq P(M)-\frac{1}{2}=P(M-1)-\frac{1}{2}+T_M<T_M,$$

d'où, en posant

$$P(M)-\frac{1}{2}=\eta T_M,$$

$$\sigma=2\{p(1-m)+(m-p)\eta\}T_M,$$

avec

$$0<\eta<1$$

En l'écrivant sous la forme

$$\sigma=2[pq+(p-m)(p-\eta)]T_M,$$

on voit que si $n\to+\infty$, m tendra vers p et $\frac{T_M}{T_R}$ tendra vers 1, tandis que T_R est, comme on sait, asymptotiquement équivalent à $\frac{1}{\sqrt{2\pi npq}}$. De sorte que σ est, comme σ', asymptotiquement équivalent à $\frac{\sqrt{2pq}}{\pi n}$. (Cette valeur asymptotique de σ' avait été calculée directement I, 1, § 11, p. 45, 46 en remplaçant asymptotiquement la loi binomiale par la seconde loi de Laplace.)

Cas de Poisson. — On peut généraliser les formules précédentes

qui concernent ce qu'on peut appeler le cas de Bernoulli : celui d'un événement dont la probabilité p reste constante au cours de n épreuves. On se placera dans le cas examiné par Poisson, celui où la probabilité de l'événement fortuit considéré E prend des valeurs quelconques données d'avance $p_1, p_2, \ldots, p_n$ dans les épreuves de rangs respectifs $1, 2, \ldots, n$. On emploiera le même raisonnement. Mais, la valeur moyenne de X_k étant cette fois p_k et son écart quadratique moyen étant $\sqrt{p_k q_k}$ (avec $q_k = 1 - p_k$), on aura

$$\bar{r} = p_1 + \ldots + p_n, \qquad \bar{f} = \frac{p_1 + \ldots + p_n}{n}; \tag{73}$$

$$\mu' = \sqrt{p_1 q_1 + \ldots + p_n q_n}, \qquad \mu = \frac{\sqrt{p_1 q_1 + \ldots + p_n q_n}}{n}; \tag{74}$$

$$\overline{(r - \bar{r})^3} = \sum_{i=1}^{i=n} p_i q_i (q_i - p_i), \tag{75}$$

$$\nu = \frac{1}{n} \sqrt[4]{\sum_{i=1}^{i=n} p_i q_i (p_i^2 + q_i^2 - p_i q_i) + 6 \sum_{i,j} p_i q_i p_j q_j} \tag{75 bis}$$

CONVERGENCE UNIFORME DE LA LOI BINOMIALE

On trouve dans tous les Traités de probabilité, la démonstration du théorème, dû à de Moivre, d'après lequel les sommes $P(s) = \sum_{r \leqq s} T_r$ de probabilités $T_r = \varpi_r^{(n)}$ de la répétition r d'un événement au cours de n épreuves peuvent s'exprimer approximativement quand n est grand, au moyen de l'intégrale de Laplace.

Nous allons reprendre la démonstration classique de ce résultat afin d'établir en même temps, l'*uniformité* [1] de la convergence, en précisant d'ailleurs, en vue du Chapitre V et aussi pour son intérêt propre, sur quel champ de variation elle a lieu.

On a

$$\varpi_r^{(n)} = \frac{n!}{r!\,(n-r)!} p^r q^{n-r}.$$

---

[1] On trouvera des indications sur ce point dans le livre de M. Paul Lévy (I, p. 229-230)

Au moyen de la formule de Stirling

$$\lim_{n \to \infty} \frac{n!}{n^n e^{-n} \sqrt{2\pi n}} = 1,$$

on transforme $\varpi_r^{(n)}$, pour en obtenir une expression asymptotique en posant

$$n! = n^n e^{-n} \sqrt{2\pi n}\, e^{\varepsilon_n}.$$

On obtient

$$\varpi_r^{(n)} = \left(\frac{np}{r}\right)^r \left(\frac{nq}{n-r}\right)^{n-r} \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{\frac{r}{n}\frac{n-r}{n}}} e^{\varepsilon_n - \varepsilon_r - \varepsilon_{n-r}}.$$

Il est naturel de faire intervenir la déviation de r à partir de sa valeur moyenne np, soit $u = r - np$. Alors, en supposant $pq \neq 0$,

$$(76) \quad \varpi_r^{(n)} = \left(1 + \frac{u}{np}\right)^{-r} \left(1 - \frac{u}{nq}\right)^{-n+r} \frac{1}{\sqrt{2\pi npq}} \frac{1}{\sqrt{\frac{r}{np}\frac{n-r}{nq}}} e^{\varepsilon_n - \varepsilon_r - \varepsilon_{n-r}}.$$

Un premier essai, que pourra faire le lecteur, montre qu'il vaut mieux, pour la simplicité du résultat, faire intervenir la déviation « réduite » $x = \frac{u}{\sqrt{2npq}}$. Faisons maintenant la supposition *que r varie avec n de sorte que x reste dans un intervalle fini fixe*. Par exemple $|x| < a$. Alors comme

$$r = np + x\sqrt{2npq}, \qquad n - r = nq - x\sqrt{2npq},$$

r et $n-r$ tendront vers l'infini *uniformément* quand n croissant indéfiniment, x varie entre $-a$ et $+a$. Dans ces conditions $\varepsilon_n - \varepsilon_r - \varepsilon_{n-1}$ et $e^{\varepsilon_n - \varepsilon_r - \varepsilon_{n-r}}$ convergeront uniformément vers zéro et vers l'unité. De même, dans l'expression (76) de $\varpi_r^{(n)}$, le terme

$$\frac{1}{\sqrt{\frac{r}{np}\frac{n-r}{nq}}} = \frac{1}{\sqrt{1 + x\sqrt{\frac{2q}{np}}}\sqrt{1 - x\sqrt{\frac{2p}{nq}}}}$$

convergera uniformément vers l'unité. Considérons enfin le terme

$$K_n = \left(1 + \frac{u}{np}\right)^{-r} \left(1 - \frac{u}{nq}\right)^{-n+r}$$

On a, en prenant le logarithme

$$\mathcal{L} K_n = -r\,\mathcal{L}\left(1 + \frac{u}{np}\right) + (r-n)\,\mathcal{L}\left(1 - \frac{u}{nq}\right).$$

Or on a

$$\mathcal{L}(1+t) = t - \frac{t^2}{2}[1+\omega(t)] \tag{77}$$

avec

$$\lim_{t \to 0} \omega(t) = 0.$$

Donc

$$\mathcal{L}K_n = -r\left[\frac{u}{np} - \frac{u^2}{2n^2p^2}(1+\omega_n)\right] + (n-r)\left[\frac{u}{nq} + \frac{u^2}{2n^2q^2}(1+\omega'_n)\right], \tag{78}$$

où, puisque

$$\left|\frac{u}{n}\right| < a\sqrt{\frac{2pq}{n}},$$

$\frac{u}{n}$ converge uniformément vers zéro, donc aussi

$$\omega_n = \omega\left(\frac{u}{np}\right), \qquad \omega'_n = \omega\left(-\frac{u}{nq}\right).$$

D'ailleurs, après réduction, (78) devient

$$\mathcal{L}K_n = -x^2 + \frac{2x^3(q-p)}{\sqrt{2npq}} + \eta_n$$

Le second terme de la somme converge uniformément vers zéro, d'autre part

$$\eta_n = pqx^2\left[\frac{r}{n}\frac{\omega_n}{p^2} + \frac{n-r}{n}\frac{\omega'_n}{q^2}\right];$$

d'où en appelant ω''_n le plus grand en valeur absolue des deux nombres ω_n, ω'_n

$$|\eta_n| < pqa^2\left(\frac{1}{p^2} + \frac{1}{q^2}\right)\omega''_n,$$

ω''_n convergeant uniformément vers zéro, il en sera de même de η_n.

On voit donc que $\mathcal{L}K_n + x^2$ [1] converge uniformément vers zéro ou que $e^{x^2}K_n$ converge uniformément vers l'unité et, par suite, il en est

---

[1] Il ne faut pas oublier que x ne peut être supposé fixe quand n croît, puisque $x = \frac{r-np}{\sqrt{2npq}}$ où r et n restent entiers. C'est seulement pour simplifier l'écriture que nous écrivons x au lieu de $x_r^{(n)}$.

de même de $e^{x^2}\varpi_r^{(n)}\sqrt{2\pi pqn}$. On peut donc d'abord écrire

$$\varpi_r^{(n)} \sim \frac{e^{-x^2}}{\sqrt{2\pi pqn}}$$

et dire que $\frac{e^{-x^2}}{\sqrt{2\pi pqn}}$ est une expression asymptotique de $\varpi_r^{(n)}$. Mais, pour bien préciser, nous résumerons le résultat acquis *en écrivant*

$$\varpi_r^{(n)} = \frac{e^{-x^2}}{\sqrt{2\pi pqn}}(1 + h_r^{(n)}), \tag{79}$$

où $h_r^{(n)}$ *converge uniformément vers zéro* quand n croît, lorsque r varie par valeurs entières (de zéro à n), *de façon qu'en posant* $r = np + x\sqrt{2pqn}$, x *reste dans un intervalle fini* indépendant de n.

Signalons à cette occasion les résultats suivants. M. S. Bernstein (1) a démontré que, pour n assez grand, $\varpi_r^{(n)}$ se rapproche le plus de $\sqrt{\frac{n}{2\pi r(n-r)}}\, e^{-x^2}$ quand r est l'entier le plus approché de

$$np + x\sqrt{2npq} + x^2\left(\frac{q-p}{3}\right).$$

MM. Mirimanoff et R. Dovaz (1) ont prouvé que

$$\varpi_r^{(n)} = \frac{e^{-x^2}}{\sqrt{2\pi npq}}\left\{1 + \frac{x(q-p)}{\sqrt{2npq}}\left(-1 + \frac{2x^2}{3}\right)\right\} + \frac{\varepsilon}{npq\sqrt{npq}}$$

avec $|\varepsilon| < 0,1$, si $npq \geqq 9$ et $|x| \leqq \sqrt{2}$.

Remarque. — On peut supposer x compris entre des limites fixes; mais puisque $x = \frac{r - np}{\sqrt{2npq}}$ où r et n sont des entiers et où n croît, il est impossible de supposer x fixe. On ne saurait donc dire en toute rigueur que $\varpi_r^{(n)}\sqrt{2\pi npq}$ tend vers e^{-x^2}. Mais si r varie avec n de sorte que x tende vers un nombre x_0 indépendant de r et de n, alors on pourra dire que ce produit tend vers $e^{-x_0^2}$.

Probabilité maximum. — Considérons en particulier *la répétition la plus probable* R. On a vu (p. 87) que $R = (n+1)p - c$ où $0 \leqq c < 1$; la valeur correspondante de x est donc $\frac{p-c}{\sqrt{2npq}}$ et tend vers

zéro avec $\frac{1}{n}$, Donc e^{-z^2} tend vers l'unité, dans ce cas et par suite $\varpi_R^{(n)}\sqrt{2\pi npq}$ tend vers l'unité. On a alors, pour le terme maximum,

$$0 \leqq \varpi_r^{(n)} \leqq \varpi_R^{(n)} \sim \frac{1}{\sqrt{2\pi npq}}.$$

Il en résulte, en particulier, que $\varpi_r^{(n)}$ *converge uniformément vers zéro* quand n croît et, cette fois, *sans aucune restriction sur la répétition* r qui peut varier librement par valeurs entières entre zéro et n. On peut même préciser en disant · $\varpi_r^{(n)}$ est, quand n croît, un infiniment petit d'ordre au moins égal à celui de $\frac{1}{\sqrt{n}}$, quelle que soit la variation de r avec n.

Cas où l'écart réduit n'est pas borné. — La formule asymptotique (79) a été obtenue en supposant x borné. Le raisonnement peut s'étendre à un cas plus général qui peut parfois être utile (p. 102 et 223). Pour le voir, introduisons un terme de plus dans le développement limité (77) de $\mathcal{L}(1+t)$. On a ainsi

$$\mathcal{L}(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3}[1+\varphi(t)]$$

avec

$$\lim_{t \to 0} \varphi(t) = 0,$$

d'où

$$\omega(t) = -\frac{2t}{3}[1+\varphi(t)]$$

Revenons aux formules de la page 91. Alors

$$\omega_n = -\frac{2}{3}\frac{u}{np}(1+\varphi_n), \qquad \omega'_n = \frac{2}{3}\frac{u}{nq}(1+\varphi'_n),$$

où φ_n et φ'_n tendent vers zéro uniformément quand $\frac{u}{n} = x\sqrt{\frac{2pq}{n}}$ tend vers zéro uniformément. On a donc

$$\eta_n = pq\,x^2\left[-\frac{r}{np^2}\frac{1+\varphi_n}{p} + \frac{n-r}{nq^2}\frac{1+\varphi'_n}{q}\right]\frac{2}{3}\frac{u}{n} = \frac{(2pq)^{\frac{3}{2}}}{3}\frac{x^3}{\sqrt{n}}\eta'_n$$

avec

$$|\eta'_n| < \frac{|1+\varphi_n|}{p^3} + \frac{|1+\varphi'_n|}{q^3}.$$

D'où

$$\mathcal{L}K_n + x^2 = \frac{x^3}{\sqrt{n}}\left\{\frac{2(q-p)}{\sqrt{2pq}} + \frac{(2pq)^{\frac{1}{2}}}{3}\eta'_n\right\}. \tag{80}$$

$\mathcal{L}K_n + x^2$ converge uniformément vers zéro quand n croît, non seulement comme ci-dessus dans le cas où x reste borné, mais aussi dans le cas plus général où $\frac{x^3}{\sqrt{n}}$ tend vers zéro uniformément. Dans ce cas, x peut devenir infini mais d'un ordre inférieur à celui de $n^{\frac{1}{6}}$.

A fortiori $\frac{x}{\sqrt{n}}$ converge uniformément vers zéro donc r, $n-r$ tendent uniformément vers l'infini (car on a, par exemple

$$\frac{r}{np} = 1 + x\sqrt{\frac{2q}{np}}\Bigg)$$

et les termes

$$e^{\varepsilon_n - \varepsilon_r - \varepsilon_{n-r}}\frac{1}{\sqrt{1+x\sqrt{\frac{2q}{np}}}\sqrt{1-x\sqrt{\frac{2p}{nq}}}}$$

convergent aussi vers 1 uniformément. En résumé, dans la formule (79), $h_r^{(n)}$ converge encore uniformément vers zéro quand $\frac{x}{\sqrt[6]{n}}$ ou $\frac{r-np}{n^{\frac{2}{3}}}$ ou encore $(f-p)n^{\frac{1}{3}}$ convergent uniformément vers zéro.

Dans le cas important où $p=q$, on peut étendre encore la validité du raisonnement en poussant le développement de $\mathcal{L}(1+x)$ un terme plus loin, on obtient $\varphi(t) = -\frac{3t}{4}[1+\psi(t)]$ avec $\lim\limits_{t \to 0}\psi(t) = 0$ et

$$\mathcal{L}K_n + x^2 = \frac{x^4}{n}\left[-\frac{1}{3} + \eta''_n\right]$$

avec

$$\eta''_n = \frac{1}{4}\left[\frac{n-r}{n}\psi'_n - \frac{r}{n}\psi_n\right], \qquad |\eta''_n| < \frac{1}{4}\{|\psi_n| + |\psi'_n|\},$$

où ψ_n, ψ'_n, et par suite η''_n convergent uniformément vers zéro quand $\frac{x}{\sqrt{n}}$ tend vers zéro. Or on voit que dans ce cas si $\frac{x^4}{n}$ ou $\frac{x}{n^{\frac{1}{4}}}$ tendent uniformément vers zéro, $(\mathcal{L}K_n + x^2)$ tendra aussi uniformément

vers zéro, etc. Finalement, on voit que si $p=q$, $h_r^{(n)}$ converge uniformément vers zéro, même si x n'est pas borné pourvu que $\frac{x}{n^{\frac{1}{4}}}$ converge uniformément vers zéro, ce qui est une hypothèse moins stricte que dans le cas où $p \neq q$.

Réciproque. — Le raisonnement montre aussi que réciproquement : si $\frac{x}{\sqrt{n}}$ tend vers zéro, $h_r^{(n)}$ ne peut tendre vers zéro *que si* $\frac{x}{\sqrt[4]{n}}$ tend aussi vers zéro quand $p=\frac{1}{2}$ et même seulement quand $\frac{x}{\sqrt[6]{n}}$ tend vers zéro quand $p \neq \frac{1}{2}$

Comme

$$\frac{x}{\sqrt{n}} = \frac{r-np}{n\sqrt{2pq}} = \left(\frac{r}{n}-p\right)\frac{1}{\sqrt{2pq}}$$

et

$$\frac{x}{\sqrt[t]{n}} = \sqrt[t]{n}\,\frac{\left(\frac{r}{n}-p\right)}{\sqrt{2pq}},$$

ceci revient à dire que : *si la fréquence* $f=\frac{r}{n}$ *d'un événement de probabilité* p, *converge vers* p *quand* n *croît indéfiniment, la condition nécessaire et suffisante pour l'équivalence asymptotique de la probabilité* $\varpi_r^{(n)}$ *de cette fréquence avec l'expression* $\frac{e^{-\frac{n(f-p)^2}{2pq}}}{\sqrt{2\pi npq}}$ *est que l'infiniment petit* $f-p$ *soit d'ordre en* $\frac{1}{n}$ *supérieur à* $\frac{1}{3}$ *si* $p \neq \frac{1}{2}$, *à* $\frac{1}{4}$ *si* $p=\frac{1}{2}$.

Qu'arrive-t-il si $f-p$ ne répond pas à ces conditions? Pour simplifier la discussion, ne nous occupons que du cas où $p \neq \frac{1}{2}$ et supposons d'abord $f-p$ infiniment petit. Si son ordre est non pas supérieur mais exactement égal à celui de $\frac{1}{\sqrt[3]{n}}$ $\left(p \text{ étant } \neq \frac{1}{2}\right)$; plus précisément *si* $(f-p)\sqrt[3]{n}$ *a une limite déterminée* $A \neq 0$, alors en employant un terme de plus dans le développement du logarithme, la méthode indiquée ci-dessus, montre que la quantité $k=\left(\frac{\varpi_r^{(n)}}{\frac{e^{-x^2}}{\sqrt{2\pi npq}}}\right)$ tend

vers e^{α} avec $\alpha = \frac{A^3(q-p)}{6p^2q^2}$. (Si $p=q$, le rapport k tend encore vers 1 comme nous le savions d'avance puisque, dans le cas actuel, $f-p$ est d'ordre supérieur à $\frac{1}{4}$.) Mais si $p \neq \frac{1}{2}$, on voit que le rapport k tend vers une quantité e^{α} *qui peut prendre toutes les valeurs positives autres que l'unité* : < 1 si $A(q-p) < 0$, > 1 si $A(q-p) > 0$. Le signe de A est celui de $f-p$ quand n est assez grand. Donc si à partir d'une valeur assez grande de n, f reste en dehors de p et q, mais du côté de p, la limite du rapport k sera < 1 ; dans le cas contraire, elle sera > 1.

Supposons maintenant que $f-p$ soit un infiniment petit d'un ordre inférieur à $\frac{1}{3}$, p restant $\neq \frac{1}{2}$. Alors $\mathcal{L}k$ tend vers l'infini et sa partie principale étant $n(f-p)^3\frac{q-p}{6p^2q^2}$, son signe sera déterminé comme plus haut. *Si*, pour n assez grand, f reste en dehors de p et q, mais du côté de p, alors k *tendra vers zéro* et, dans le cas contraire, k *tendra vers l'infini*.

Enfin, reste le cas où f ne tend pas vers p. Si par exemple f tend vers une limite $\varphi \neq p$, les développements de $\log k$ ne sont plus valables. Mais alors on peut écrire

$$k = \frac{\varpi_r^{(n)}}{\dfrac{e^{-y^2}}{\sqrt{2\pi npq}}} = \left[\frac{C_n^r\,\varphi^r(1-\varphi)^{n-r}}{\dfrac{e^{-y^2}}{\sqrt{2\pi npq}}}\right]\left[\left(\frac{p}{\varphi}\right)^r\left(\frac{q}{1-\varphi}\right)^{n-r} e^{y^2-y'^2}\right] = hl$$

avec

$$y = \frac{(r-n\varphi)}{\sqrt{2npq}}.$$

Le second crochet l, comme on le voit facilement, ne peut tendre que vers o ou $+\infty$, quand φ n'a pas une valeur particulière. Le premier crochet h, où $\frac{r}{n} \to \varphi$ s'étudiera comme on avait étudié k quand $\frac{r}{n} \to p$.

On voit donc que l'on aura des résultats différents suivant que f converge plus ou moins étroitement vers φ. Nous renvoyons pour la suite de la discussion à une petite note qui sera consacrée à cet objet.

Probabilité que l'écart réduit soit compris entre deux limites

données. — Évaluons de même la probabilité $P_n(\lambda, \lambda')$ pour que

$$(81) \qquad \lambda\sqrt{2npq} < r - np < \lambda'\sqrt{2npq}.$$

Soient ρ, ρ', les bornes entières de r vérifiant cette double inégalité. On a

$$P_n(\lambda, \lambda') = \sum_{\rho}^{\rho'} \omega_r^{(n)} = \sum_{\rho}^{\rho'} \frac{e^{-x_r^2}}{\sqrt{2\pi npq}}(1 + h_r^{(n)}),$$

en précisant par la notation x_r la valeur de x qui correspond à r. Or, soit $H^{(n)}$, la plus grande des valeurs absolues des $h_r^{(n)}$ pour $\rho \leqq r \leqq \rho'$: il est clair que la dernière égalité peut s'écrire

$$P_n(\lambda, \lambda') = S_n(1 + h^{(n)})$$

avec

$$|h^{(n)}| < H^{(n)}$$

et

$$S_n = \sum_{\rho}^{\rho'} \frac{e^{-x_r^2}}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{\pi}} \sum_{r=\rho}^{r=\rho'} e^{-x_r^2}(x_{r+1} - x_r)$$

Si λ et λ' varient avec n, mais restent bornés, par exemple compris dans $(-a, +a)$, $H^{(n)}$ qui dépend de λ et de λ', convergera, d'après ce qui précède, uniformément vers zéro. Si λ, λ' restent fixes, il est clair que S_n aura pour limite l'intégrale

$$J(\lambda, \lambda') = \frac{1}{\sqrt{\pi}} \int_{\lambda}^{\lambda'} e^{-x^2}\, dx$$

On a donc d'abord, *pour* λ, λ' *fixes, l'égalité classique*

$$(82) \qquad \lim_{n \to \infty} P_n(\lambda, \lambda') = J(\lambda, \lambda'),$$

c'est-à-dire,

$$(82\ bis) \quad \lim_{n \to \infty} \Pr.\left\{\lambda\sqrt{2npq} < r - np < \lambda'\sqrt{2npq}\right\} = \frac{1}{\sqrt{\pi}} \int_{\lambda}^{\lambda'} e^{-t^2}\, dt$$

pour λ, λ', p donnés.

Signalons le résultat suivant de M. S. Bernstein (1) : l'inégalité

$$\left| r - np - \frac{\lambda^2}{3}(q - p) \right| < \lambda\sqrt{2npq}$$

considérée, à une unité près, a la probabilité

$$\frac{2}{\sqrt{\pi}}\int_0^{\lambda} e^{-x^2}\,dx$$

à $2e^{-\sqrt[3]{2npq}}$ près, si

$$0 < \lambda < \sqrt[6]{2npq} \qquad \text{et} \qquad npq \geqq 365$$

Uniformité de la convergence. — L'égalité (82) perd son sens quand λ, λ' ne sont pas fixes. Par contre, on peut essayer d'étendre à ce cas, l'égalité

$$\lim_{n \to \infty}[\mathrm{P}_n(\lambda, \lambda') - \mathrm{J}(\lambda, \lambda')] = 0.$$

Il reste encore à établir l'uniformité de la convergence de cette formule par rapport à λ et λ'.

I. Supposons encore $-a < \lambda < \lambda' < a$, pour commencer, et observons que

$$\int_{x_r}^{x_{r+1}} e^{-x^2}\,dx = e^{-x_r'^2}(x_{r+1} - x_r) = e^{-x_r^2}(x_{r+1} - x_r)(1 + \beta_r^{(n)}),$$

où x'_r est un certain nombre compris entre x_r et x_{r+1} et où

$$1 + \beta_r^{(n)} = e^{x_r^2 - x_r'^2}$$

Or on a

$$|x_r^2 - x_r'^2| < \frac{2a}{\sqrt{2pqn}},$$

de sorte que $\beta_r^{(n)}$ converge vers zéro uniformément quand, n croissant, r varie dans les limites fixées. Si $\mathrm{B}^{(n)}$ est la plus grande des valeurs absolues de $\beta_r^{(n)}$ quand

$$|r - np| < a\sqrt{2npq},$$

$\mathrm{B}^{(n)}$ tend vers zéro avec $\frac{1}{n}$ et l'on a

$$\int_{x_\rho}^{x_{\rho'+1}} e^{-x^2}\,dx = \sum_{\rho}^{\rho'} e^{-r_r^2}(x_{r+1} - x_r)(1 + \beta_r^{(n)}) = \sqrt{\pi}\,\mathrm{S}_n(1 + b^{(n)})$$

avec

$$|b^{(n)}| < \mathrm{B}^{(n)}.$$

D'ailleurs

$$\int_{x_\rho}^{x_{\rho'+1}} e^{-x^2}\,dx = \int_{\lambda}^{\lambda'} e^{-x^2}\,dx - \int_{\lambda}^{x_\rho} e^{-x^2}\,dx + \int_{\lambda'}^{x_{\rho'+1}} e^{-x^2}\,dx$$
$$= \sqrt{\pi}\,[\,J(\lambda, \lambda') + c^{(n)}\,]$$

où

$$\sqrt{\pi}\, c^{(n)} \leqq \frac{2}{\sqrt{2pqn}}.$$

De sorte que $c^{(n)}$ converge uniformément vers zéro ; et l'on a

$$P_n(\lambda, \lambda') - J(\lambda, \lambda') = [\,J(\lambda, \lambda') + c^{(n)}\,]\frac{[1 + h^{(n)}]}{1 + b^{(n)}} - J(\lambda, \lambda')$$
$$= J(\lambda, \lambda')\left[\frac{1 + h^{(n)}}{1 + b^{(n)}} - 1\right] + c^{(n)}\left[\frac{1 + h^{(n)}}{1 + b^{(n)}}\right].$$

Puisque $0 \leqq J(\lambda, \lambda') \leqq 1$, il en résulte que $P_n(\lambda, \lambda') - J(\lambda, \lambda')$ converge vers zéro avec $\frac{1}{n}$ et cela uniformément (comme $h^{(n)}$, $b^{(n)}$ et $c^{(n)}$) quand λ, λ' varient (en même temps que n) dans l'intervalle $(-a, +a)$

II. Considérons maintenant le cas où λ, λ' varient arbitrairement entre $-\infty$ et $+\infty$ en même temps que n croît et donnons-nous un nombre positif α arbitraire On peut d'abord choisir a de sorte que $1 - J(-a, +a) < \frac{\alpha}{10}$, par exemple Puis on peut déterminer N, tel que l'on ait, par exemple,

$$(83) \qquad |\,P_n(\lambda, \lambda') - J(\lambda, \lambda')\,| < \frac{\alpha}{10} \qquad \text{pour } n > N,$$

quels que soient λ, λ' dans $-a$, $+a$. On aura en particulier

$$1 - P_n(-a, a) = [\,1 - J(-a, +a)\,] + [\,J(-a, +a) - P_n(-a, +a)\,] < \frac{\alpha}{5}.$$

Or pour $\lambda < \lambda'$, on a six cas à distinguer :

$$-a \leqq \lambda \leqq \lambda' \leqq a$$
$$P_n(\lambda, \lambda') - J(\lambda, \lambda') = [\,P_n(\lambda, \lambda') - J(\lambda, \lambda')\,],$$
$$\lambda \leqq -a \leqq \lambda' \leqq a$$
$$= [\,P_n(-a, \lambda') - J(-a, \lambda')\,] + P_n(\lambda, -a) - J(\lambda, -a);$$
$$\lambda \leqq -a < a \leqq \lambda'$$
$$= [\,P_n(-a, a) - J(-a, a)\,]$$
$$+ P_n(\lambda, -a) - J(\lambda, -a) + P_n(a, \lambda') - J(a, \lambda');$$
$$\lambda \leqq \lambda' \leqq -a < a :$$
$$= P_n(\lambda, \lambda') - J(\lambda, \lambda'),$$

....

Dans chaque cas, le terme entre crochets est $< \frac{\alpha}{5}$ en vertu de (83) et chacun des autres termes est $< \frac{\alpha}{5}$ comme étant inférieur à $1 - P_n(-a, a)$ ou à $1 - J(-a, a)$. On a donc, en définitive, quels que soient λ et λ', $(\lambda < \lambda')$

$$|P_n(\lambda, \lambda') - J(\lambda\ \lambda')| < \alpha \qquad \text{pour } n \quad N.$$

le nombre N étant indépendant de λ et de λ'

En résumé, on peut dire que *l'erreur commise en remplaçant la probabilité* $P_n(\lambda, \lambda')$ (pour que la répétition soit contenue entre $np + \lambda\sqrt{2npq}$ et $np + \lambda'\sqrt{2npq}$) *par l'intégrale*

$$J(\lambda, \lambda') = \frac{1}{\sqrt{\pi}} \int_{\lambda}^{\lambda'} e^{-t^2}\, dt,$$

converge vers zéro avec $\frac{1}{n}$ *et cela uniformément quand* $\lambda, \lambda', (\lambda < \lambda')$ *varient avec* n, arbitrairement et indépendamment de n, *de* $-\infty$ *à* $+\infty$.

Voir aussi Note C (Addition 8γ), p. 295.

Équivalence asymptotique. — On ne pourrait en déduire sans précaution qu'il y a équivalence asymptotique uniforme de $P_n(\lambda, \lambda')$ et de $J(\lambda, \lambda')$, c'est-à-dire que $\frac{P_n(\lambda, \lambda')}{J(\lambda, \lambda')}$ converge uniformément vers l'unité quand n croît si λ, λ' varient arbitrairement.

En effet, si la différence $\lambda' - \lambda$ pouvait être aussi petite qu'on veut avec $\frac{1}{n}$, le quotient $\frac{P_n}{J}$ pourrait tendre aussi bien vers l'infini ou vers zéro, par exemple, que vers 1. Car, si par exemple, on prenait (R étant la répétition la plus probable dans n épreuves)

$$\lambda = \frac{\frac{1}{3} + R - np}{\sqrt{2npq}}, \qquad \lambda' = \frac{\frac{2}{3} + R - np}{\sqrt{2npq}},$$

$P_n(\lambda, \lambda')$ serait nul, $J(\lambda, \lambda') \neq 0$ et par suite $\frac{P_n}{J}$ tendrait vers zéro.

Si, au contraire, on prenait

$$\lambda = \frac{-\beta_n + R - np}{\sqrt{2npq}} \qquad \lambda' = \frac{\beta_n + R - np}{\sqrt{2npq}}$$

avec $0 < \beta_n < 1$, on aurait $P_n(\lambda, \lambda') = \varpi_R^{(n)}$ et

$$J(\lambda, \lambda') = \frac{1}{\sqrt{\pi}} \int_\lambda^{\lambda'} e^{-x^2} dx = \frac{2\beta_n}{\sqrt{2\pi pqn}} e^{-\lambda''^2} \qquad (\lambda < \lambda'' < \lambda'),$$

d'ou

$$\frac{P_n(\lambda, \lambda')}{J(\lambda, \lambda')} = \left[\sqrt{2\pi pqn}\, \varpi_R^{(n)}\right] \frac{1}{[2\beta_n e^{-\lambda'^2}]}.$$

Le premier crochet tend vers 1, le second crochet est inférieur à $2\beta_n$ qu'on peut faire tendre vers zéro. Dans ce cas. $\frac{P_n}{J}$ tend vers l'infini.

Par contre, si l'on fait varier λ, λ', $(\lambda < \lambda')$, entre $-\infty$ et $+\infty$ en ne les assujettissant qu'a la condition

$$J(\lambda, \lambda') \geqq \varepsilon,$$

où ε est un nombre positif arbitraire indépendant de n, alors. on aura

$$\frac{P_n(\lambda, \lambda')}{J(\lambda, \lambda')} = 1 + g^{(n)}$$

avec

$$|g^{(n)}| = \left|\frac{P_n - J}{J}\right| \leqq \frac{|P_n - J|}{\varepsilon},$$

de sorte que $g^{(n)}$ convergera uniformément vers zéro et $\frac{P_n}{J}$ vers l'unité. La condition ci-dessus sera réalisée, par exemple, s'il existe deux nombres positifs β et γ indépendants de n, tels que

$$\lambda' - \lambda > \beta \qquad \text{et} \qquad -\gamma < \lambda < \lambda' < \gamma$$

Cas où $\int_\lambda^{\lambda'} e^{-x^2} dx$ tend vers zéro. — Cependant, on peut établir l'équivalence même dans des cas où $J(\lambda, \lambda')$ tend vers zéro avec $\frac{1}{n}$. Supposons par exemple que λ et λ' varient avec n de sorte que le nombre des termes correspondants de $P_n(\lambda, \lambda')$ croisse indéfiniment Comme on a

$$(\rho' - \rho) < (\lambda' - \lambda)\sqrt{2npq} \leqq \rho' - \rho + 2,$$

c'est le cas où $(\lambda' - \lambda)\sqrt{n}$ croît indéfiniment.

Si λ croît indéfiniment avec n et $\lambda' > \lambda$, il est clair qu'alors $J(\lambda, \lambda')$ tend vers zéro. Mais nous n'avons pas besoin de cette hypothèse, il

nous suffira de supposer par exemple, $\lambda \geqq 0$. Posons

$$\lambda_r = \lambda + \frac{(r-\rho)}{\sqrt{2npq}}$$

et

$$\varpi_r^{(n)} = \frac{e^{-x_r^2}}{\sqrt{2npq\pi}}[1 + h_r^{(n)}].$$

On a vu que $h_r^{(n)}$ converge uniformément vers zéro si, en posant

$$x_r = \frac{r - np}{\sqrt{2npq}},$$

$\frac{x_r}{\sqrt[6]{n}}$ converge uniformément vers zéro. Mais pour $\rho \leqq r \leqq \rho'$, on a

$$\lambda < x_r < \lambda'.$$

Supposons maintenant que $\frac{\lambda'}{\sqrt[6]{n}}$ converge uniformément vers zéro, alors en supposant $0 \leqq \lambda \leqq \lambda'$; $\frac{x_r}{\sqrt[6]{n}}$ et $\frac{\lambda}{\sqrt[6]{n}}$ convergeront de même.

On a

$$P_n(\lambda, \lambda') = \sum_{r=\rho}^{r=\rho'} \frac{e^{-x_r^2}}{\sqrt{2npq\pi}}[1 + h_r^{(n)}].$$

Or

$$\int_{\lambda_r}^{\lambda_{r+1}} e^{-u^2}\,du = \frac{1}{\sqrt{2npq}} e^{-x_r'^2} \qquad \text{avec } \lambda_r < x_r' < \lambda_{r+1}$$

Donc

$$P_n(\lambda, \lambda') = \sum_{r=\rho}^{r=\rho'} \left[\frac{1}{\sqrt{\pi}} \int_{\lambda_r}^{\lambda_{r+1}} e^{-u^2}\,du\right] e^{x_r'^2 - x_r^2}[1 + h_r^{(n)}].$$

Or on a

$$\rho - np - \theta = \lambda\sqrt{2npq} \qquad \text{avec } 0 < \theta \leqq 1.$$

D'où

$$x_r = \lambda_r + \frac{\theta}{\sqrt{2npq}};$$

ainsi, pour n assez grand, x_r et x_r' appartiennent à l'intervalle λ_r, λ_{r+1}; on aura donc

$$|x_r'^2 - x_r^2| \leqq 2(\lambda_{r+1} - \lambda_r)\lambda_{\rho'+1} \leqq \frac{2}{\sqrt{2npq}}\left(\lambda' + \frac{1}{\sqrt{2npq}}\right).$$

Puisque $\frac{\lambda}{\sqrt{n}}$ converge uniformément vers zéro, il en sera de même de $x_i'^2 - x_i^2$ et l'on pourra poser

$$e^{x_i'^2 - x_i^2} = 1 + k_i^{(n)},$$

où $k_i^{(n)}$ converge uniformément vers zéro. Alors

$$\mathrm{P}_n(\lambda, \lambda') = \sum_{i=\rho}^{i=\rho'} \left[\frac{1}{\sqrt{\pi}} \int_{\lambda_i}^{\lambda_{i+1}} e^{-u^2}\, du\right] [1 + k_i^{(n)}][1 + h_i^{(n)}],$$

d'où

$$\mathrm{P}_n(\lambda, \lambda') = \left[\frac{1}{\sqrt{\pi}} \int_{\lambda_\rho}^{\lambda_{\rho'+1}} e^{-u^2}\, du\right] (1 + l_n),$$

où l_n converge uniformément vers zéro. Or

$$\int_{\lambda_\rho}^{\lambda_{\rho'+1}} e^{-u^2}\, du = \int_{\lambda}^{\lambda'} e^{-u^2}\, du + \int_{\lambda'}^{\lambda_{\rho'+1}} e^{-u^2}\, du.$$

Comme on a supposé $\lambda \geqq 0$, on aura

$$\int_{\lambda}^{\lambda'} e^{-u^2}\, du > (\lambda' - \lambda)\, e^{-\lambda'^2},$$

$$\left|\int_{\lambda'}^{\lambda_{\rho'+1}} e^{-u^2}\, du\right| = |\lambda_{\rho'+1} - \lambda'|\, e^{-\lambda''^2},$$

où λ'' est entre λ' et $\lambda_{\rho'+1}$ et

$$\lambda' - \frac{1}{\sqrt{2npq}} \leqq \lambda_{\rho'+1} \leqq \lambda' + \frac{1}{\sqrt{2npq}}.$$

Alors

$$(84) \qquad \frac{\left|\int_{\lambda'}^{\lambda_{\rho'+1}} e^{-u^2}\, du\right|}{\int_{\lambda}^{\lambda'} e^{-u^2}\, du} \qquad \frac{e^{\lambda'^2 - \lambda''^2}}{\sqrt{2npq}\,(\lambda' - \lambda)}$$

avec

$$|\lambda'^2 - \lambda''^2| \leqq \frac{1}{\sqrt{2npq}}\left(2\lambda' + \frac{1}{\sqrt{2npq}}\right).$$

Puisque $\frac{\lambda'}{\sqrt{n}}$ converge uniformément vers zéro, il en sera de même de $(\lambda'^2 - \lambda''^2)$ et puisqu'en outre $\sqrt{n}(\lambda' - \lambda)$ converge uniformément vers l'infini, le rapport (84) convergera uniformément vers zéro. On

aura donc

$$\int_{\lambda_\rho}^{\lambda'_{\rho'+1}} e^{-u^2}\,du = \left[\int_{\lambda}^{\lambda'} e^{-u^2}\,du\right](1 + u_n),$$

où u_n converge uniformément vers zéro; et par suite

$$P_n(\lambda, \lambda') = J(\lambda, \lambda')(1 + u_n)(1 + l_n)$$

En résumé, *en supposant seulement* que $\frac{\lambda'}{\sqrt[6]{n}}$ et $\frac{1}{(\lambda' - \lambda)\sqrt{n}}$ convergent uniformément vers zéro et que $\lambda' > \lambda > 0$, on a l'équivalence asymptotique

$$P_n(\lambda, \lambda') \sim J(\lambda, \lambda')$$

et, plus précisément, *si l'on pose*

$$\frac{P_n(\lambda, \lambda')}{J(\lambda, \lambda')} = 1 + \gamma_n,$$

γ_n converge uniformément vers zéro avec $\frac{1}{n}$.

Examinons maintenant le cas où $\lambda' = +\infty$ (ou $\lambda = -\infty$), c'est-à dire, — par exemple, pour $\lambda' = +\infty$ —, comparons la probabilité $Q_n(\lambda)$ que $r - np > \lambda\sqrt{2npq}$ à l'intégrale

$$K(\lambda) = \frac{1}{\sqrt{\pi}}\int_{\lambda}^{+\infty} e^{-x^2}\,dx$$

On verra encore que $\frac{Q_n(\lambda)}{K(\lambda)}$ converge uniformément vers 1 lorsque λ varie de façon quelconque quand n croît pourvu que $K(\lambda)$ reste supérieur à un nombre positif fixe, ce qui revient à dire que λ reste inférieur à un nombre fixe quand n croît. Comme cette condition est simplement suffisante, il est bon de s'assurer que le rapport $\frac{Q_n(\lambda)}{K(\lambda)}$ ne converge pas toujours vers l'unité quand λ n'est soumis à aucune restriction. Il suffit de prendre, pour chaque valeur de n, λ tel que $np + \lambda\sqrt{2pqn} > n$; alors $r > n$, donc $Q_n(\lambda) = 0$, $K(\lambda) > 0$ et le rapport $\frac{Q_n(\lambda)}{K(\lambda)}$ tend vers zéro avec $\frac{1}{n}$.

Cependant il n'est pas sans intérêt pour la suite de noter que le rapport $\frac{Q_n(\lambda)}{K(\lambda)}$ peut encore converger uniformément vers l'unité quand n tend vers $+\infty$ sans supposer λ borné, pourvu que la varia-

tion de λ avec n satisfasse à une condition simple : celle même qui assure la convergence uniforme de $\varpi_j^{(n)} e^{\tau^2} \sqrt{2\pi npq}$. C'est ce que nous verrons p. 222.

SECTION IV · HISTORIQUE

Fonctions génératrices, fonctions caractéristiques. — Nous faisons grand usage, dans ce qui précède et dans ce qui suit, de la fonction des probabilités totales $C(x)$ d'une variable aléatoire X

Elle fournit en effet une représentation claire de la distribution des probabilités des diverses valeurs de X. Elle n'a été systématiquement employée que depuis une époque assez récente. M. von Mises a beaucoup contribué à en faire apprécier la commodité.

En renonçant au caractère intuitif de cette fonction, on peut utiliser d'autres fonctions représentatives qui ont l'avantage de faire correspondre à l'opération $Z = X + Y$ une transformation, plus simpel que celle fournie par la formule (64), des fonctions représentatives de X et de Y dans celle de Z. Nous visons ici les fonctions génératrices et les fonctions caractéristiques

Dans l'œuvre immense qu'il a consacrée à la Théorie des probabilités, Laplace attachait une importance particulière à la notion de « fonction génératrice ». Cette notion lui a été utile, d'une part, pour la résolution des équations aux différences. D'autre part, dit-il luimême, pour « déterminer avec facilité les limites de la probabilité des résultats et des causes, indiqués par les événements considérés en grand nombre, et les lois suivant lesquelles cette probabilité approche de ses limites à mesure que les événements x se multiplient ». Dans la période moderne, la méthode des fonctions génératrices semble être tombée en discrédit. Mais en même temps, et précisément en vue du second problème signalé par Laplace, était fait de plus en plus usage de la notion de fonction caractéristique, employée par Poincaré, et dont l'invention était attribuée à Cauchy.

Il est d'abord facile de se rendre compte que, si, pour une variable aléatoire donnée, X, sa fonction génératrice et sa fonction caractéristique sont bien deux fonctions distinctes, du moins leur introduction est basée sur la même idée. Il s'agit de remplacer dans un grand nombre de raisonnements, la considération mal commode de la suite double des valeurs x_j prises par X et de leurs probabilités corres-

pondantes p_j par la considération plus simple d'une fonction ordinaire d'une variable auxiliaire.

Dans ses applications aux jeux de hasard, ou au calcul des différences, Laplace avait surtout affaire à des variables aléatoires ne prenant que des valeurs entières $x_j = j$. C'est ce qui l'a conduit à appeler fonction génératrice de X la somme

$$\Phi(u) = \sum_j p_j u^j,$$

c'est-à-dire la valeur moyenne de u^X

$$\Phi(u) = \mathfrak{M}(u^X),$$

où u est un nombre arbitraire certain.

Mais, comme il pouvait être parfois plus commode d'aboutir à une fonction finie quelle que soit la variable aléatoire X, Laplace a aussi envisagé une autre fonction de u obtenue comme valeur moyenne d'une fonction de X et de u autre que u^X. Il a ainsi utilisé une fonction qui est la moyenne de e^{itX}, c'est-à-dire (t, X étant réels) la somme de la valeur moyenne de $\cos tX$ et du produit de i par la valeur moyenne de $\sin tX$. Ces deux valeurs moyennes étant nécessairement comprises entre -1 et $+1$, $\mathfrak{M}e^{itX}$ est une fonction $\varphi(t)$ qui est nécessairement finie et même bornée.

En s'affranchissant de la restriction de Laplace, consistant ici à ne considérer que les variables aléatoires ne prenant que des valeurs entières, on voit qu'il s'agit bien ici tout simplement de la fonction dite maintenant caractéristique. Il est d'ailleurs clair que si l'on admet pour u des valeurs complexes, on a $\varphi(t) = \Phi(e^{it})$ de sorte que la distinction entre fonction caractéristique et fonction génératrice est surtout formelle. Il faut en outre noter (comme l'a montré M. Molina, à qui nous devons ces renseignements historiques et d'autres encore) que Laplace s'est parfaitement rendu compte des deux avantages reconnus de nos jours aux fonctions caractéristiques.

D'une part, la fonction caractéristique d'une somme de deux variables aléatoires indépendantes est le produit de leurs fonctions caractéristiques puisque

$$\mathfrak{M}[e^{it(X+Y)}] = \mathfrak{M}(e^{itX}e^{itY}) = [\mathfrak{M}(e^{itX})][\mathfrak{M}(e^{itY})].$$

Laplace n'ayant pas donné un nom à la fonction caractéristique ne pouvait énoncer explicitement ce théorème; mais il en a fait clairement usage (*Œuvres*, Livre II, Chap. IV, p. 309; Chap. II). La fonction génératrice jouit également de la même propriété, puisque

$$\mathfrak{M}(u^{X+Y}) = \mathfrak{M}(u^X u^Y) = \mathfrak{M}(u^X)\,\mathfrak{M}(u^Y)$$

et ceci justifie également l'usage fréquent que Laplace a fait de cette fonction. Cette propriété appartient d'ailleurs également à toute fonction de u définie sous la forme $\mathfrak{M}\{[\alpha(u)]^X\}$, quelle que soit la fonction $\alpha(u)$.

D'autre part, on sait que la distribution des probabilités des diverses valeurs d'une variable aléatoire peut se déduire de la connaissance de sa fonction caractéristique (ce Traité, I, 2, § 13, p. 26). Cette propriété précieuse était aussi connue de Laplace (*Œuvres*, p. 83, 84) chez qui l'on trouve (sous une notation différente) les relations réciproques

$$(85)\qquad \left\{\begin{aligned} \Phi(t) &= \sum_{n=0}^{n=\infty} p_n e^{int}, \\ p_n &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \Phi(t)\, e^{-int}\, dt \end{aligned}\right.$$

C'est sans doute pour ces raisons que M. von Mises appelle la fonction caractéristique : *la transformée de Laplace*

Par contre, on doit à Cauchy, d'avoir étendu l'exemple de la fonction caractéristique au cas général, cas plus difficile où la variable aléatoire ne prend pas seulement des valeurs entières, ni même des valeurs quelconques, en nombre fini, et où l'intégrale de Fourier vient remplacer la simple sommation qui figure dans (85) Depuis lors, les travaux de M. Paul Lévy et de M. von Mises ont montré la grande fécondité de la notion de fonction caractéristique. On les trouvera résumés dans ce Traité, I, 2, p. 24.

On doit d'ailleurs observer que la fonction des probabilités totales jouit d'un avantage analogue à la première propriété qui vient d'être attribuée à la fonction caractéristique d'une somme.

Considérons n variables indépendantes $X_1, \ldots, X_n$ définies sur la même catégorie d'épreuves et désignons par Y_n la plus grande, par $A_j(x)$ et $G_n(x)$ les fonctions des probabilités totales de X_j et de Y_n. Il est clair que $G_n(x)$ est la probabilité pour que $X_1, \ldots, X_n$ soient

simultanément $< x$. On a donc

$$G_n(x) = A_1(x)A_2(x) \quad A_n(x).$$

Dans le cas où l'on ne considère que deux variables aléatoires X_1 et X_2, on aura la formule simple

$$G_2(x) = A_1(x)A_2(x)$$

Celle-ci montre que *les deux opérations* : faire la somme de deux variables aléatoires, déterminer la plus grande de ces deux variables, *ont respectivement le même effet* la première sur les fonctions caractéristiques, la seconde sur les fonctions de probabilités totales.

Enfin observons qu'on peut considérer aussi la fonction des probabilité totales $C(x)$ de X comme une valeur moyenne, à savoir $C(x) = \mathfrak{M}\alpha(X - x)$ où $\alpha(u) = 1$ pour $u < 0$ et $\alpha(u) = 0$ pour $u \geq 0$.

Digression sur la loi des erreurs d'observation. — Nous avons utilisé cette propriété de la fonction des probabilités totales pour étudier la loi de probabilité de l'écart maximum (Fréchet, **115**), et en tirer des conséquences (Fréchet, **114**) quant à la portée véritable des démonstrations modernes de la loi des erreurs d'observation. Celles-ci sont, en effet, toutes basées sur l'hypothèse que l'erreur résultante est la somme (ou au moins, une combinaison linéaire) des erreurs partielles. Nous avons montré qu'en suivant exactement l'une des méthodes les plus modernes, mais en partant d'une autre hypothèse (bonne ou non, ceci n'étant pas en cause pour notre but), on obtient une loi réduite limite autre que la « seconde loi de Laplace » (*voir* paragraphe suivant).

Ainsi était mis en lumière le rôle fondamental joué dans les « démonstrations » de cette loi par l'hypothese de l'additivité des erreurs partielles. Or, cette hypothèse, qui est une des plus plausibles est, pourtant, trop simplificatrice et sûrement éloignée de la réalité qui est plus complexe.

Encore un peu d'histoire. — L'intégrale $J(\lambda, \lambda')$ rencontrée plus haut se retrouve aussi dans la théorie des erreurs. Mais, contrairement à une opinion si répandue que l'on parle couramment de l'intégrale de Gauss et de la loi des erreurs de Gauss, c'est à Laplace que l'on doit cette introduction. Bertrand avait contribué à cette légende en

écrivant, au sujet de la loi des erreurs d'observation : « Euler, Bernoulli, Lagrange et Laplace ont fait des hypothèses démenties par les faits et mal justifiées par des preuves sans vraisemblance.

» Gauss, plus heureux, a déduit d'un raisonnement fort simple une loi que la démonstration laisserait douteuse, mais que les conséquences justifient. »

En réalité, Gauss lui-même n'a pas été satisfait de sa première démonstration et en a proposé une seconde; ses deux démonstrations sont aujourd'hui abandonnées et les démonstrations les plus modernes sont basées sur une conception profonde de Laplace.

M. Castelnuovo dans son remarquable ouvrage (1, vol. II) s'exprime ainsi à la p. 148 . « Parmi les conceptions grandioses et ardues de la « Théorie analytique des probabilités de Laplace », si riche de résultats et de méthodes, il y a une question sur laquelle l'illustre mathématicien semble avoir médité trente années, de la position du problème en 1780 (2) jusqu'à la formule résolutoire publié dans un mémoire de 1810 (3, 4, 5).

» Laplace doit avoir eu l'intuition, dès ses premieres recherches sur le calcul des probabilités, que la raison du succes de celui-ci dans des problemes variés de philosophie naturelle réside dans le fait suivant la résultante d'un grand nombre de quantités qui varient au hasard suivant des lois inconnues obéit, dans ses oscillations autour de la valeur moyenne, a une loi simple et uniforme. »

Quant à l'affirmation de Bertrand que la loi dite de Gauss est justifiée par ses conséquences alors que les hypotheses de Laplace sont démenties par les faits, il suffit pour être en droit de la rejeter, de se reporter aux deux lois d'erreurs proposées successivement par Laplace La seconde étant identique à la loi dite de Gauss, les faits ne peuvent que les justifier ou les démentir a la fois toutes les deux. Quant à la premiere — où le rôle de e^{-x^2} est tenu par $e^{-|x|}$, — si Laplace ne s'y est pas tenu, il faut toutefois observer que des vérifications expérimentales récentes de M. E. B. Wilson (1) ont montré qu'elle se conformait aux faits presque aussi bien que la seconde. Elle est, d'autre part, d'une application pratique plus simple.

Reste la question de date. M. E.-B. Wilson (1) a proposé d'employer les expressions : *première et seconde loi d'erreur* et de cesser de désigner la seconde sous les termes de loi de Gauss ou de loi normale. Pour justifier sa proposition (à laquelle nous nous sommes, depuis,

conformés), il rappelle que ces deux lois d'erreurs ont été publiées par Laplace en 1774 (1) et en 1778 (2). Et il fait observer qu'en dépit de la précocité bien connue de Gauss, il est peu probable que ce dernier ait déjà fait sa découverte avant d'avoir deux ans. Ceci restitue à sa juste limite (mais ne doit pas conduire à nier) l'importance de la contribution de Gauss à la théorie des erreurs.

En ce qui concerne l'intégrale

$$J(\lambda, \lambda') = \frac{1}{\sqrt{\pi}} \int_{\lambda}^{\lambda'} e^{-v^2} dv,$$

nous venons de rappeler que c'est à Laplace qu'on doit sa considération dans la théorie des erreurs et plus généralement dans l'étude asymptotique de la loi de probabilité d'une somme de variables aléatoires indépendantes. Mais il ne faut pas oublier, comme l'a fait observer M. Karl Pearson (1), que c'est à de Moivre (1, 2) qu'est due l'introduction de cette intégrale pour représenter approximativement les probabilités des épreuves répétées dans le cas de Bernoulli.

C'est d'ailleurs à Kramp (1) qu'est due (sur l'instigation de Laplace) la première table numérique de cette intégrale.

CHAPITRE IV.

L'INÉGALITÉ DE BIENAYMÉ ET SES GÉNÉRALISATIONS.

L'inégalité de Bienaymé. — En 1853, Bienaymé (1, p 167) à qui la science est redevable de nombreux travaux malheureusement méconnus, a fait connaître une formule remarquable à la fois par sa simplicité et son utilité et par la facilité de sa démonstration. Nous en étudierons plus loin diverses généralisations. Nous nous contenterons d'abord de l'établir et d'en déduire une démonstration extrêmement simple du théorème de Bernoulli.

Soit μ, l'écart quadratique moyen de deux variables aléatoires X, $q(t)$, la probabilité pour que $|X - Y| < t$.

On a

$$\mu^2 = \int_0^{+\infty} t^2 \, d\, q(t).$$

Soit ε un nombre positif arbitraire, on a évidemment

$$\mu^2 \geqq \int_\varepsilon^{+\infty} t^2 \, d\, q(t) \geqq \varepsilon^2 [1 - q(\varepsilon)]$$

D'où l'inégalité

$$q(\varepsilon) \geqq 1 - \frac{\mu^2}{\varepsilon^2}. \tag{86}$$

Ou encore, en appelant $P(t)$ la probabilité pour que $|X - Y| \geqq t\mu$

$$P(t) \leqq \frac{1}{t^2}. \tag{87}$$

Telle est la formule de Bienaymé particulièrement remarquable en ce qu'elle fournit une limite supérieure de $P(t)$ qui est indépendante de

la fonction $q(t)$, et par suite de la loi de probabilité à laquelle obéit X.

En particulier, si l'on prend pour Y un nombre certain x, on voit que : la probabilité pour que X diffère de x d'au moins t fois l'écart quadratique moyen de X avec x est au plus égale à $\frac{1}{t^2}$. Comme cet écart moyen est minimum quand $x = \overline{X}$, on voit que la formule obtenue est la plus avantageuse dans ce dernier cas. C'est celui qui a été considéré par Bienaymé et par la plupart des auteurs, mais il peut être utile aussi d'envisager le cas général, particulièrement quand on ne connaît pas la valeur moyenne de X (*voir* p. 116).

On attribue souvent à Tchebichef [*voir*, par exemple (I, 1, p. 146. 147)] un théorème qu'on énonce sous la forme suivante :

Soit $X = X_1 + X_2 + \ldots$ la somme de plusieurs variables aléatoires indépendantes. La probabilité pour que X soit compris entre

$$\overline{X_1} + \overline{X_2} + \ldots - t\sqrt{\overline{X_1^2} + \overline{X_2^2} + \ldots - (\overline{X_1})^2 - (\overline{X_2})^2}$$

et

$$\overline{X_1} + \overline{X_2} + \ldots + t\sqrt{\overline{X_1^2} + \overline{X_2^2} + \ldots - (\overline{X_1})^2 - (\overline{X_2})^2}$$

est toujours au moins égale à $1 - \frac{1}{t^2}$.

En réalité, ce n'est là que la conséquence immédiate, écrite sous une forme compliquée, des deux formules déjà signalées (61). (87) énoncées toutes deux par Bienaymé. Quand, en effet, $Y = \overline{X}$, la probabilité $1 - P(t)$ n'est autre que la probabilité pour que X soit compris entre $\overline{X} - t\mu$ et $\overline{X} + t\mu$.

Or, d'apres (61), avec une autre notation

$$\mu^2 = \mu_1^2 + \mu_2^2 + \ldots$$

avec

$$\mu_1^2 = \overline{X_1^2} - (\overline{X_1})^2, \qquad \mu_2^2 = \overline{X_2^2} - (\overline{X_2})^2, \qquad \ldots$$

Donc $1 - P(t)$ est la probabilité considérée dans ce théorème et d'après (87), elle est $\geqq 1 - \frac{1}{t^2}$.

Démonstration du théorème de Bernoulli. — Dans le premier fascicule de cet ouvrage (I, 1), le théorème de Bernoulli a été d'abord

démontré à la page 41 en faisant usage de la formule d'approximation de la loi binomiale par l'intégrale de Laplace Θ. On peut arriver au même résultat en quelques mots, comme l'a signalé, le premier, Tchebichef (1, p. 183), en faisant usage de la formule de Bienaymé. C'est la seconde démonstration, donnée page 149 du présent Traité (I, 1) et que nous reproduisons (pour en montrer la simplicité) en la dégageant du théorème dit de Tchebichef cité ci-dessus. En effet, d'après la formule (86), la probabilité pour que la fréquence $f = \frac{r}{n}$ d'un événement de probabilité constante p diffère de p de moins de ε en valeur absolue est au moins égale à $1 - \frac{\mu^2}{\varepsilon^2}$. Or, ici $\mu^2 = \frac{pq}{n}$. Par suite, cette probabilité, comprise entre 1 et $1 - \frac{1}{n}\frac{pq}{\varepsilon^2}$ tend vers l'unité quand le nombre n des épreuves croît indéfiniment.

Démonstration du théorème de Poisson — Dans le cas de Poisson, la valeur moyenne de la fréquence n'est pas, en général, indépendante du nombre des épreuves. Nous avons vu (p. 89) qu'elle est égale à la moyenne arithmétique

$$p^{(n)} = \frac{p_1 + \quad + p_n}{n}$$

des probabilités de E relatives aux épreuves considérées. Nous ne pouvons plus dire que cette fréquence $f^{(n)}$ diffère probablement peu d'un nombre certain indépendant de n. Mais l'écart quadratique moyen de $p^{(n)}$ et $f^{(n)}$, soit $\mu^{(n)} = \sqrt{\frac{p_1 q_1 + \quad + p_n q_n}{n}} \leqq \frac{1}{\sqrt{n}}$, tend encore vers zéro avec $\frac{1}{n}$. La probabilité que $|p^{(n)} - f^{(n)}| < \varepsilon$ étant encore comprise entre 1 et $1 - \left[\frac{\mu^{(n)}}{\varepsilon}\right]^2$ tendra donc encore vers l'unité.

Extension de ces théorèmes. — Plus généralement, soient X_1, X_2, ..., les valeurs aléatoires prises par une même variable aléatoire X au cours de n épreuves indépendantes, et v_n la moyenne arithmétique des n valeurs X_1, ..., X_n. On a vu plus haut, page 72, que si X a une valeur moyenne déterminée et finie $\overline{X}$ et si X a avec $\overline{X}$ un écart quadratique moyen μ fini, l'écart quadra-

tique moyen λ de v_n avec $\overline{X}$ est égal à $\frac{\mu}{\sqrt{n}}$. Par suite, λ tend vers zéro avec $\frac{1}{n}$ et, ici encore, la probabilité que v_n diffère de $\overline{X}$ de moins de ε en valeur absolue tend vers l'unité quand n croît indéfiniment.

Plus généralement encore, considérons le cas où $X_1, X_2, \ldots,$ sont des variables aléatoires indépendantes quelconques, ayant chacune, X_k, une valeur moyenne $\overline{X_k}$ et un écart quadratique moyen μ_k finis et déterminés.

Si les μ_k ont une même borne supérieure A, on aura, en vertu de la formule de Bienaymé (61), pour l'écart quadratique moyen λ de v_n et $\overline{v_n}$,

$$\lambda^2 \leqq \frac{A^2}{n}$$

La formule de Bienaymé montre encore que l'inégalité $|v_n - \overline{v_n}| < \varepsilon$ est un événement dont la probabilité tend vers l'unité quand n croît indéfiniment. Il en est de même quand les μ_k, sans être bornés dans leur ensemble, sont tels que $\mu_1^2 + \ldots + \mu_n^2$ soit, quand n croît, un infiniment grand d'ordre inférieur à celui de n^2.

Un cas particulier important est celui où les X_k sont bornés dans leur ensemble. Alors les $\overline{X_k}$ et les μ_k existent, sont finis et bornés dans leur ensemble et la probabilité que

$$\left| \frac{X_1 + \ldots + X_n}{n} - \frac{\overline{X_1} + \ldots + \overline{X_n}}{n} \right| < \varepsilon$$

tend encore vers l'unité quand n croît indéfiniment.

Nous verrons, plus loin, au Chapitre V, p. 254, qu'on peut notablement préciser plusieurs de ces énoncés, en introduisant de nouveaux modes de convergence.

Première généralisation de l'inégalité de Bienaymé. — La démonstration de la formule de Bienaymé s'applique évidemment au cas plus général où, au lieu de considérer un écart d'ordre deux, on considère un écart d'ordre positif r quelconque entier ou non. Soient λ_r l'écart moyen d'ordre r de deux variables aléatoires X, Y et ε un nombre positif arbitraire. On a

$$(\lambda_r)^r = \int_0^{+\infty} u^r\, dq(u) \geqq \varepsilon^r \int_\varepsilon^{\infty} dq(u) = \varepsilon^r [1 - q(\varepsilon)];$$

d'où

$$q(\varepsilon) \geqq 1 - \left(\frac{\lambda_r}{\varepsilon}\right)^r. \tag{88}$$

Ou encore, en appelant $\varpi(u)$ la probabilité que $|X - Y|$ soit au moins égal à u fois l'écart moyen λ_r

$$\varpi(u) \leqq \frac{1}{u^r} \tag{89}$$

Pour $r = 2$, on obtient la formule de Bienaymé. La forme la plus simple de la formule (88) s'obtient évidemment pour $r = 1$. Nous appellerons la formule correspondante

$$q(\varepsilon) \geqq 1 - \frac{\sigma}{\varepsilon},$$

ou σ est l'écart moyen de X avec Y, formule de Markoff. Les extensions au cas ou r est un entier pair, puis au cas ou r est un nombre positif arbitraire ont été signalées successivement par M Medolaghi (1) et par M. Cantelli (1)

Il est intéressant de rappeler ici une remarque de M. Ragnar Frisch (1), d'après laquelle la formule de Bienaymé est la plus avantageuse de son espece qu'il soit possible de donner. Afin de bien montrer que ce résultat, tres intéressant, dépend essentiellement du sens qu'on donne aux mots « de son espèce », nous montrerons que la formule (89) d'ordre r est, elle aussi, en un certain sens. la plus avantageuse. Pour cela, observons que la probabilité $\varpi(u)$ pour que $|X - Y| \geqq u\lambda_r$ est $\leqq \frac{1}{u^r}$. Supposons qu'il existe une fonction $F(u)$ telle que, pour tout couple de variables aléatoires X, Y, d'écart moyen λ_r, on ait

$$\varpi(u) \leqq F(u) \leqq \frac{1}{u^r}$$

au moins pour les valeurs de u où la formule de Markoff est utile, c'est-à-dire pour $u > 1$. Il s'agirait de démontrer qu'on aura nécessairement alors $F(u) \equiv \frac{1}{u^r}$ pour $u \geqq 1$. Mais la démonstration de ce fait n'est pas plus simple que celle d'un fait plus général qui sera établi plus loin page 119 et auquel il nous suffit de renvoyer le lecteur.

On voit alors le paradoxe : pour chaque valeur de r la formule

d'ordre r serait la plus avantageuse. Il ne serait donc pas vrai que la formule de Bienaymé occupât la place privilégiée que lui assigne M. Frisch. En réalité, le paradoxe disparaît, si l'on met la formule d'ordre r (89) sous la forme équivalente (88). Sous cette forme, on va montrer (p. 119) que la formule d'ordre r est la plus avantageuse *parmi celles* où l'on suppose fini *et connu* l'écart moyen d'ordre r, λ_r, et où la formule n'est utilisée que pour les valeurs de ε où elle n'est pas illusoire, c'est-à-dire pour $\varepsilon > \lambda_r$.

Ainsi, par exemple, la formule de Bienaymé (86) est la plus avantageuse *parmi* celles qui font usage de l'existence et de la connaissance de l'écart quadratique moyen de X et Y et seulement pour les valeurs de $\varepsilon > \mu$. Mais, de même, la formule de Markoff est aussi la plus avantageuse parmi celles qui font usage de l'écart moyen σ et seulement pour $\varepsilon > \sigma$.

Cette comparaison avait déjà fait l'objet d'une étude beaucoup plus générale due à Tchebicheff (2, 3). Celui-ci cherche la borne inférieure de $\int_\alpha^\beta f(x)\,dx$, quand, α, β étant donnés sur un segment donné (a, b), on prend pour $f(x)$ toutes les fonctions telles que les intégrales $\int_a^b x^n f(x)\,dx$ $(n = 1, 2, \ldots)$ aient des valeurs données d'avance. Sa méthode est fondée sur l'emploi des « résidus intégraux ». Des résultats de Tchebicheff, M. Cantelli (1) avait déduit (antérieurement à la note de M. Frisch) que plusieurs de ses formules sont les plus avantageuses de leur espèce.

Remarque. — La plupart des auteurs traitant des écarts moyens considéraient seulement les écarts moyens d'ordre entier pair à partir de la moyenne. Nous avons déjà vu qu'on pouvait généraliser au cas des écarts moyens d'ordre non entier et qu'on pouvait aussi prendre l'écart moyen de X avec un nombre certain différent de $\overline{X}$ et même avec un autre nombre aléatoire Y. Donnons un exemple de l'utilité de cette seconde généralisation. Nous allons, grâce à elle, obtenir très simplement, et d'une manière à la fois plus générale et aussi précise, la solution d'un problème résolu par M. Cantelli (1, p. 16, 18), savoir : trouver une limite inférieure de la probabilité P pour qu'une variable aléatoire X soit comprise entre deux nombres certains donnés α et β.

Prenons pour Y le nombre certain $\frac{\alpha + \beta}{2}$. P est aussi la probabilité

pour que

$$|X - Y| < \frac{\beta - \alpha}{2}.$$

D'après la formule de Bienaymé (86)

$$P \geqq 1 - \frac{\overline{(X - Y)^2}}{\left(\frac{\beta - \alpha}{2}\right)^2}.$$

D'ailleurs, si μ est l'écart quadratique moyen de X avec sa valeur moyenne $\overline{X}$, on a

$$\overline{(X - Y)^2} = \mu^2 + \left(\overline{X} - \frac{\alpha + \beta}{2}\right)^2;$$

d'où

$$(90) \qquad P \geqq 1 - \frac{\mu^2 + \left(\frac{\alpha + \beta}{2} - \overline{X}\right)^2}{\left(\frac{\beta - \alpha}{2}\right)^2}.$$

D'ailleurs pour que cette formule donne un résultat, il faut que

$$\left(\frac{\beta - \alpha}{2}\right)^2 \quad \mu^2 + \left(\frac{\alpha + \beta}{2} - X\right)^2.$$

ou

$$|\beta - \overline{X}|\,|\overline{X} - \alpha| \geqq \mu^2,$$

ce qui exige d'abord que les valeurs α et β encadrent la valeur moyenne $\overline{X}$ et ensuite qu'elles en soient assez éloignées

M. Cantelli donne deux formules, pour exprimer P dans le seul cas particulier où α et β sont de la forme

$$\alpha = \overline{X} - \left(t + \frac{2}{t}\right)\mu \qquad \beta = \overline{X} + t\mu,$$

c'est-à-dire où β, seul, étant pris arbitrairement $(> \overline{X})$, α est pris égal à

$$\overline{X} - \left[\beta - \overline{X} + \frac{2\mu^2}{\beta - \overline{X}}\right].$$

Il trouve alors

$$P \geqq \frac{t^2}{t^2 + 1} \quad [\text{form } (e), \S\ 9, \text{p. } 16]$$

et

$$P \geqq \frac{t^2}{t^2 + 1} - \frac{t^2}{t^2 + (t^2 + 2)^2} \quad [\text{form. } (f), \S\ 9\ \text{p. } 18].$$

Nous pouvons négliger la seconde, moins précise que la première, et comparer la premiere avec le résultat (90) déduit directement de la formule de Bienaymé. Or dans le cas particulier actuel, on constate que notre formule (90) redonne exactement la première formule de M. Cantelli (2)

$$P \geqq \frac{t^2}{t^2+1}.$$

Nouvelle généralisation de la formule de Bienayme. — La formule d'ordre r (88) peut aussi s'écrire

$$(91) \qquad p(\varepsilon) \leqq \frac{\overline{|X-Y|^r}}{\varepsilon^r} \leqq \frac{\overline{A(|X-Y|)}}{A(\varepsilon)},$$

en posant $A(x) = x^r$ Au lieu de faire intervenir pour A, une fonction particulière de $|X-Y|$, on peut utiliser une fonction plus générale de x, définie pour $x \geqq 0$ et former, avec M Slutsky, « le A-moment de $|X-Y|$ », c'est-à-dire la valeur moyenne m_A de $A(|X-Y|)$.

Pour obtenir la formule plus générale annoncée, nous nous placerons avec M. Cantelli dans le cas où la fonction $A(x)$ est $\geqq 0$ pour $x \geqq 0$ et où ε étant une valeur > 0 donnée, on a

$$A(t) \geqq A(\varepsilon) > 0 \qquad \text{pour } t \geqq \varepsilon$$

Alors

$$m_A = \int_0^{+\infty} A(t)\, dq(t) \geqq A(\varepsilon)\, p(\varepsilon),$$

d'où

$$(92) \qquad \boxed{p(\varepsilon) \leqq \frac{m_A}{A(\varepsilon)}.}$$

On peut prendre, en particulier, $A(x) \equiv x^r$; dans ce cas la formule ci-dessus se réduit à la formule (88) d'ordre r.

Les hypothèses faites sur $A(\varepsilon)$ peuvent être rendues plus strictes. (ce qui sera plus commode pour la suite) sans qu'il en résulte d'inconvénient.

On peut toujours former une fonction $A_2(t)$ qui soit pour $t \geqq 0$: non décroissante et au plus égale à $A(t)$, qui soit continue quand $A(t)$ est continue, qui soit nulle pour $t = 0$ et égale à $A(\varepsilon)$ pour $t = \varepsilon$ [1].

---

[1] On peut, par exemple, appeler d'abord $A_1(t)$ la borne inférieure de $A(t')$

Il est clair que $m_{A_2} \leqq m_A$ et, comme $A_2(\varepsilon) = A(\varepsilon)$, on voit que la formule obtenue en substituant A_2 à A

$$p(\varepsilon) \leqq \frac{m_{A_2}}{A_2(\varepsilon)}$$

sera au moins aussi avantageuse.

Or la fonction $A_2(t)$ est une fonction qui, pour toute valeur $\geqq 0$ de t, est non décroissante et qui est nulle pour $t = 0$. On voit donc qu'on ne perd rien à se limiter à de telles fonctions. C'est ce que fait M. Slutsky (1, p. 34) qui a obtenu indépendamment la même formule (1, p. 41). M. Cantelli qui, le premier, a obtenu cette formule (92) (1, p. 9), d'ailleurs sous une forme un peu différente, l'a écrite, sans s'y arrêter, ne l'utilisant que comme simple intermédiaire, pour obtenir une formule que nous signalerons plus loin (p. 123). Pourtant comme nous allons le montrer, cette formule (92) est, en un certain sens, *la plus avantageuse de son espèce*. Plus précisément, si, pour une fonction A donnée, il existe une fonction $H(m_A, \varepsilon) \geqq 0$ telle que pour tout couple de variables aléatoires X, Y, on ait

$$p(\varepsilon) \leqq H(m_A, \varepsilon) \leqq \frac{m_A}{A(\varepsilon)},$$

m_A étant le A-moment de $|X - Y|$, alors on a nécessairement

$$H(m_A, \varepsilon) = \frac{m_A}{A(\varepsilon)}$$

pour toutes les valeurs ε où la formule (92) est utile, c'est-à-dire pour ε tel que $A(\varepsilon) > m_A$.

En effet, prenons pour X et Y deux valeurs aléatoires telles que $|X - Y|$ ne prenne que deux valeurs : $\varepsilon > 0$ avec la probabilité k, 0 avec la probabilité $1 - k$. On aura

$$m_A = k\, A(\varepsilon),$$

---

pour $t' \geqq t \geqq 0$. C'est évidemment, pour $t \geqq 0$, une fonction $\geqq 0$, non décroissante, au plus égale à $A(t)$. De plus, puisqu'on suppose $A(t) \geqq A(\varepsilon) > 0$ pour $t \geqq \varepsilon$, on aura évidemment $A_1(\varepsilon) = A(\varepsilon)$. Et $A_1(t)$ sera continue, si $A(t)$ est continue. Il suffit maintenant d'appeler $A_2(t)$ la plus petite des quantités $A_1(t)$, $\frac{t}{\varepsilon} A(\varepsilon)$, pour $t \geqq 0$, pour conserver toutes ces propriétés à la fonction A_2 et y ajouter la propriété $A_2(0) = 0$.

car nous avons vu qu'on pouvait sans inconvénient, supposer $A(o) = o$. Et, puisque k est une probabilité, on peut supposer

$$o < k < 1,$$

donc on aura bien

$$o \leq m_A \leq A(\varepsilon).$$

Alors

$$p(\varepsilon) = \Pr.\{|X - Y| \geqq \varepsilon\} = k,$$

et, d'après (92),

$$\frac{m_A}{A(\varepsilon)} = p(\varepsilon) \leqq H(m_A, \varepsilon) \leqq \frac{m_A}{A(\varepsilon)},$$

ce qui prouve la proposition. Nous nous sommes déjà servi (p. 115) du cas particulier où $A(x) \equiv x^r$ et où par suite $m_A = (\lambda_r)^r$.

Pour obtenir une formule générale encore plus semblable à celle qui concerne ce cas particulier, nous allons généraliser la notion d'écart moyen.

Écart moyen relativement à une fonction donnée. — Soit $f(x)$ une fonction continue, croissante pour $x \geqq o$ et nulle pour $x = o$. Soit d'autre part $F(z)$ sa fonction inverse, en sorte que $F[f(x)] \equiv x$. Nous appellerons écart moyen de deux variables aléatoires X, Y, relativement à la fonction $f(x)$ ou plus brievement « *f-écart moyen* de X *et de* Y » le nombre λ_f qui est bien déterminé par la relation

$$f(\lambda_f) = \overline{f(|X - Y|)},$$

qui donne

$$\lambda_f = F\left[\overline{f(|X - Y|)}\right]$$

(Si la valeur moyenne au second membre n'était pas finie, λ_f serait encore bien déterminé mais égal à $+\infty$.)

Nous allons montrer que cet écart moyen jouit d'un certain nombre de propriétés qui sont précisément celles qui ont fait le succès de l'écart moyen d'ordre r.

1° Si $\lambda_f = o$, la valeur moyenne de $f(|X - Y|)$ doit être nulle. Comme cette expression est une variable aléatoire $\geqq o$, ceci n'est possible que s'il y a une probabilité égale à 1 pour qu'elle soit nulle.

Or $f(|X - Y|)$ ne peut être nul que si X = Y. Dès lors *le « f-écart moyen » de* X *et de* Y *ne peut être nul que si* X *et* Y *sont « presque toujours égaux »*. *Réciproquement*, il est clair que dans ce dernier cas. $f(|X - Z|)$ est presque toujours nul et par suite $\lambda_f = o$.

2° Si $|X - Y|$ est un nombre certain ω, on voit immédiatement que $\lambda_f = \omega$.

3° Si $|X - Y|$ reste, dans toutes les épreuves, *compris* entre deux nombres certains ω_1 et ω_2, $\overline{f(|X - Y|)}$ sera compris entre entre $f(\omega_1)$ et $f(\omega_2)$ et par suite λ_f sera compris entre ω_1 et ω_2.

4° Si μ_f est le « f-écart moyen » de deux autres variables aléatoires Z et T, et si, dans toute épreuve, on a $|X - Y| \leqq |Z - T|$, alors $\lambda_f \leqq \mu_f$.

Ces propriétés appartiennent en particulier aux écarts moyens d'ordre r, qu'on obtient en prenant $f(x) \equiv x^r$

Lorsque $f(x)$ est une fonction bornée pour $x \geqq 0$, $f(|X - Y|)$ reste aussi inférieur dans toute épreuve à un nombre fixe et par suite sa valeur moyenne existe certainement. On voit que, dans ce cas, le « f-écart moyen » de X et de Y présente sur l'écart moyen d'ordre r de X et de Y, ce très grand avantage d'avoir une valeur finie et déterminée pour chaque couple arbitraire de valeurs aléatoires X et Y.

Nous verrons, plus loin (p. 211-214), d'autres avantages de ce nouvel écart moyen.

Dans le cas particulier où $f(x)$, dont on a vu qu'on pouvait la supposer nulle à l'origine et non décroissante pour $x \geqq 0$, est en outre continue, la formule de Cantelli-Slutsky (92) peut se mettre sous la forme suivante, qui se rapproche plus de celle de la formule d'ordre r (88).

$$p(\varepsilon) \leqq \frac{f(\lambda_f)}{f(\varepsilon)}. \tag{93}$$

Cette formule n'est d'ailleurs utile que pour $\varepsilon > \lambda_f$.

Parmi les conditions imposées à $f(x)$, il y en a une qui n'est nullement essentielle, qui ne joue qu'un rôle de commodité; c'est la condition $f(0) = 0$. Il est clair, en effet, que si $f(0) \neq 0$, la fonction $\varphi(x) = f(x) - f(0)$, satisfera à toutes les conditions imposées ci-dessus et que la relation

$$\varphi(\lambda_\varphi) = \overline{\varphi(|X - Y|)},$$

c'est-à-dire

$$f(\lambda_\varphi) - f(0) = \overline{f(|X - Y|)} - f(0),$$

définit un « φ-écart moyen » exactement égal au « f-écart moyen ».

Par contre, on pourrait aussi supposer que $f(x)$ n'est définie, continue et croissante que pour $x > 0$. On pourrait encore définir un « f-écart moyen » comme ci-dessus. Mais ici f n'étant plus nécessairement $\geqq 0$, ni bornée pour x voisin de zéro, la valeur moyenne de $f(|X - Y|)$ pourrait n'être ni finie ni infinie, mais indéterminée. C'est pourquoi nous aurions laissé ce cas de côté si nous n'avions pas eu à le rattacher à une définition donnée dans son livre, par M. Paul Lévy (1, p. 159). Cet auteur définit sous le nom d'écart moyen d'ordre zéro la limite de l'écart moyen d'ordre r quand r tend vers zéro. Et il montre que le logarithme de ce nouvel écart est égal à la valeur moyenne du logarithme de $|X - Y|$. On voit qu'en définitive l'écart moyen d'ordre zéro au sens de M. Paul Lévy n'est autre que l'écart moyen relativement à la fonction $\log x$.

Cas où l'on connaît deux écarts moyens. — Supposons maintenant que les écarts moyens de X et de Y existent et soient connus pour deux ordres distincts r et R (on peut appeler R le plus grand des deux).

Alors, on a

$$q(\varepsilon) > 1 - \left(\frac{\lambda_r}{\varepsilon}\right)^r,$$

$$q(\varepsilon) > 1 - \left(\frac{\lambda_R}{\varepsilon}\right)^R.$$

On sait que $\lambda_r < \lambda_R$. La première formule sera plus avantageuse que la seconde pour les valeurs de ε telles que

$$\varepsilon > \lambda_r \qquad \text{et} \qquad 1 - \left(\frac{\lambda_r}{\varepsilon}\right)^r > 1 - \left(\frac{\lambda_R}{\varepsilon}\right)^R$$

ou

$$\lambda_r < \varepsilon < \sqrt[R-r]{\frac{(\lambda_R)^R}{(\lambda_r)^r}}.$$

D'ailleurs, on a

$$\lambda_r < \lambda_R < \sqrt[R-r]{\frac{(\lambda_R)^R}{(\lambda_r)^r}}.$$

Donc, *la formule de l'ordre r le plus petit est celle qui convient le mieux pour les plus petites valeurs de* ε, à savoir ici pour

$$\lambda_r \leqq \varepsilon \leqq \sqrt[R-r]{\frac{(\lambda_R)^R}{(\lambda_r)^r}}.$$

La formule de l'ordre le plus grand, R, est celle qui convient le mieux pour les plus grandes valeurs de ε, à savoir ici pour

$$\varepsilon \geqq \sqrt[R-r]{\frac{(\lambda_R)^R}{(\lambda_r)^r}}.$$

En ce qui concerne les cas de $r=1$, $R=2$, quand on ne connaît que les écarts $\lambda_1=\sigma$, $\lambda_2=\mu$, la formule de Markoff convient pour

$$\sigma \leqq \varepsilon \leqq \frac{\mu^2}{\sigma},$$

celle de Bienaymé pour

$$\varepsilon \geqq \frac{\mu^2}{\sigma}.$$

Inégalités de Cantelli. — I. Mais quand on connaît deux écarts (d'ordres r et R), le raisonnement de M. Frisch ne s'applique plus et l'on peut se demander s'il n'existe pas de formule plus avantageuse dans ce cas que celle qui résulte de l'emploi des deux formules d'ordres r et R. Nous allons donner en exemple de ce fait une formule due à M. Cantelli

Considérons à cet effet, une fonction $A(\xi, k)$ satisfaisant aux conditions suivantes : c'est une fonction continue de ξ, définie et $\geqq 0$ pour $\xi \geqq 0$ et k convenablement choisi et telle que

$$A(\xi, k) \geqq A(\varepsilon, k) > 0 \qquad \text{pour} \qquad \xi > \varepsilon,$$

ε étant un nombre positif donné. On a, alors, comme on l'a vu p. 118,

$$(93\ bis) \qquad 1-q(\varepsilon) \leqq \frac{\overline{A(|X-\bar{X}|, k)}}{A(\varepsilon, k)} = M(\varepsilon, k).$$

Si $m(\varepsilon)$ est la borne inférieure de $M(\varepsilon, k)$ lorsque k varie dans le domaine choisi, on aura la formule très générale

$$(94) \qquad 1-q(\varepsilon) \leqq m(\varepsilon).$$

On voit facilement que si l'on prend $A(\xi, k)=\xi^r+k^2$, on retrouve la formule (88) d'ordre r. Mais si l'on prend $A(\xi, k)=(\xi^r+k)^2$, A satisfera aux conditions imposées si k reste supérieur à $-\varepsilon^r$ Or

$$M(\varepsilon, k)=\frac{(\lambda_{2r})^{2r}+2k(\lambda_r)^r+k^2}{(\varepsilon^r+k)^2}.$$

En posant $v = \frac{1}{\varepsilon^r + k}$, on aura

$$M(\varepsilon, k) = v^2\left\{[\varepsilon^r - (\lambda_r)^r]^2 + (\lambda_{2r})^{2r} - (\lambda_r)^{2r}\right\} + 2v\left\{(\lambda_r)^r - \varepsilon^r\right\} + 1,$$

où v pourra varier de o à $+\infty$

Le coefficient de v^2 est positif puisque $\lambda_r \leqq \lambda_{2r}$. Si $\varepsilon \geqq \lambda_r$ le coefficient de v est aussi $\geqq$ o et la borne inférieure de $M(\varepsilon, k)$ quand $v > 0$ est 1. Dans ce cas la formule (93^{bis}) ne nous apprend rien.

Si $\varepsilon > \lambda_r$, le minimum de $M(\varepsilon, k)$ quand v varie, a lieu pour la valeur de v

$$v_0 = \frac{\varepsilon^r - (\lambda_r)^r}{[\varepsilon^r - (\lambda_r)^r]^2 + (\lambda_{2r})^{2r} - (\lambda_r)^{2r}}$$

qui, étant positive, est admissible. On a donc, en substituant dans $M(\varepsilon, k)$ cette valeur de v,

$$m(\varepsilon) = \frac{(\lambda_{2r})^{2r} - (\lambda_r)^{2r}}{[\varepsilon^r - (\lambda_r)^r]^2 + [(\lambda_{2r})^{2r} - (\lambda_r)^{2r}]} < 1.$$

Finalement, on voit que, si l'on connaît λ_r et λ_{2r}, au lieu d'employer le système de formules d'ordres r et $2r$

$$(\text{I}) \begin{cases} q(\varepsilon) \geqq 1 - \left(\dfrac{\lambda_r}{\varepsilon}\right)^r, & \text{pour } \lambda_r \leqq \varepsilon \leqq \dfrac{(\lambda_{2r})^2}{\lambda_r}, \\ q(\varepsilon) \geqq 1 - \left(\dfrac{\lambda_{2r}}{\varepsilon}\right)^{2r}, & \text{pour } \dfrac{(\lambda_{2r})^2}{\lambda_r} \leqq \varepsilon, \end{cases}$$

on peut employer la formule unique, due à M. Cantelli,

$$q(\varepsilon) \geqq 1 - \frac{(\lambda_{2r})^{2r} - (\lambda_r)^{2r}}{[\varepsilon^r - (\lambda_r)^r]^2 + [(\lambda_{2r})^{2r} - (\lambda_r)^{2r}]} \qquad \text{pour } \varepsilon \geqq \lambda_r.$$

D'ailleurs, l'avantage n'est à la formule de M. Cantelli que pour $\varepsilon > \frac{(\lambda_{2r})^2}{\lambda_r}$, comme on le vérifie facilement. De sorte que, finalement, le système (I) ne doit être remplacé, pour obtenir une plus grande précision, que par le système

$$(95) \quad \boxed{(\text{II}) \begin{cases} q(\varepsilon) \geqq 1 - \dfrac{(\lambda_{2r})^{2r} - (\lambda_r)^{2r}}{[\varepsilon^r - (\lambda_r)^r]^2 + [(\lambda_{2r})^{2r} - (\lambda_r)^{2r}]} & \text{pour } \varepsilon \geqq \dfrac{(\lambda_{2r})^2}{\lambda_r}, \\ q(\varepsilon) \geqq 1 - \left(\dfrac{\lambda_r}{\varepsilon}\right)^r & \text{pour } \lambda_r \leqq \varepsilon \leqq \dfrac{(\lambda_{2r})^2}{\lambda_r}. \end{cases}}$$

ou, en posant $M_r = \overline{|X - Y|^r} = (\lambda_r)^r$,

$$[P_r\{|X-Y| \geqq \varepsilon\}] \leqq \begin{cases} \dfrac{M_r}{\varepsilon^r} & \text{pour } M_r \leqq \varepsilon^r \leqq \dfrac{M_{2r}}{M_r}, \\ \dfrac{M_{2r} - M_r^2}{[\varepsilon^r - M_r]^2 + M_{2r} - M_r^2} & \text{pour } \varepsilon^r \geqq \dfrac{M_{2r}}{M_r}. \end{cases}$$

C'est le système obtenu par M. Cantelli (7).

En particulier, quand on connaît les écarts moyens σ et μ d'ordre 1 et 2 de X et de Y, on pourra employer les inégalités

$$(96) \qquad \boxed{\begin{array}{ll} q(\varepsilon) \geqq 1 - \dfrac{\sigma}{\varepsilon} & \text{pour } \sigma \leqq \varepsilon \leqq \dfrac{\mu^2}{\sigma}, \\ q(\varepsilon) \geqq 1 - \dfrac{\mu^2 - \sigma^2}{(\varepsilon - \sigma)^2 + \mu^2 - \sigma^2} & \text{pour } \dfrac{\mu^2}{\sigma} \leqq \varepsilon \end{array}}$$

Il peut être parfois utile d'obtenir la probabilité $C(\varepsilon)$ pour que $X - Y < \varepsilon$. Par exemple si X et Y sont le doit et l'avoir d'une entreprise d'assurance, il n'y a aucun inconvénient, au contraire, à ce que le bénéfice soit très grand, tandis que l'on désire ne dépasser que rarement un certain montant pour les pertes.

Il est d'ailleurs clair que

$$q(\varepsilon) \leqq C(\varepsilon).$$

Reprenons la fonction A employée page 123, mais en la supposant positive pour tout ξ. On aura

$$\overline{A(X-Y, k)} = \int_{-\infty}^{+\infty} A(\xi, k)\, dC(\xi) \geqq \int_{\varepsilon}^{+\infty} \ldots \geqq A(\varepsilon, k)[1 - C(\varepsilon)],$$

d'où

$$1 - C(\varepsilon) \leqq \frac{\overline{A(X-Y, k)}}{A(\varepsilon, k)} = N(\varepsilon, k)$$

et, par suite,

$$1 - C(\varepsilon) \leqq n(\varepsilon),$$

$n(\varepsilon)$ étant la borne inférieure de $N(\xi, k)$ dans le domaine de variation choisi pour k.

Prenons

$$A(\xi, k) = (\xi^r + k)^2$$

en supposant cette fois r *entier* pour que ξ^r ait un sens quand $\xi < 0$

et $\xi' + k > 0$ pour que $A(\xi, k) \geqq A(\xi, k)$ quand $\xi \geqq \varepsilon > 0$. Alors

$$N(\varepsilon, k) = \frac{M_{2r} + k\mathfrak{M}_r + k^2}{(\varepsilon' + k)^2}$$
$$= v^2\{[\varepsilon' - \mathfrak{M}_r]^2 + [M_{2r} - \mathfrak{M}_r^2]\} + 2v[\mathfrak{M}_r - \varepsilon'] + 1,$$

en posant encore $v = \frac{1}{\varepsilon' + k}$, mais $\mathfrak{M}_r = \overline{(X - Y)^r}$, de sorte que $|\mathfrak{M}_r| \leqq M_r$. Le coefficient de v^2 est encore positif, car

$$M_{2r} - \mathfrak{M}_r^2 \geqq 0,$$

ce dont on s'assure en observant que le trinome en t

$$M_{2r} + 2k\mathfrak{M}_r + k^2 = \int_{-\infty}^{+\infty} (\xi' + k)^2 \, dC(\xi)$$

est toujours $\geqq 0$. Cette fois encore v peut prendre toute valeur finie positive. Le minimum $n(\varepsilon)$ de $N(\varepsilon, k)$ sera donc 1 si $\varepsilon' \leqq M_r$ et donné, si $\varepsilon' \geqq M_r$, par une formule analogue à la formule précédemment trouvée. On a donc la formule, également due à M. Cantelli,

$$(97) \quad C(\varepsilon) \geqq 1 - \frac{M_{2r} - \mathfrak{M}_r^2}{[\varepsilon' - \mathfrak{M}_r]^2 + [M_{2r} - \mathfrak{M}_r^2]} \qquad \text{pour } r \text{ entier et } \varepsilon' > \mathfrak{M}_r$$

En particulier, pour $r = 1$, si l'on prend pour Y la valeur moyenne de X, on a $\mathfrak{M}_r = \mathfrak{M}_1 = 0$ et l'on obtient [1], pour $\varepsilon > 0$,

$$(97^{bis}) \qquad \Pr\{X - \overline{X} \geqq \varepsilon\} = 1 - C(\varepsilon) \leqq \frac{\mu^2}{\varepsilon^2 + \mu^2}.$$

Une méthode générale. — M. von Mises nous a fait observer qu'on peut obtenir par une méthode uniforme un grand nombre des inégalités données plus haut.

---

(1) On aurait pu aussi borner $C(\varepsilon)$ au moyen des formules (96) ci-dessus et de l'inégalité $C(\varepsilon) \geqq q(\varepsilon)$. Il y aura bénéfice à employer la formule directe (97^{bis}) si l'on a

$$\frac{\mu^2}{\varepsilon^2 + \mu^2} < \frac{\sigma}{\varepsilon} \qquad \text{pour} \qquad \sigma \leqq \varepsilon < \frac{\mu^2}{\sigma}$$

et

$$\frac{\mu^2}{\varepsilon^2 + \mu^2} < \frac{\mu^2 - \sigma^2}{(\varepsilon - \sigma)^2 + \mu^2 - \sigma^2} \qquad \text{pour} \qquad \varepsilon > \frac{\mu^2}{\sigma}.$$

Ceci n'a pas lieu nécessairement; par exemple, la dernière inégalité n'a pas lieu si ε est assez grand, c'est-à-dire, plus précisément, si

$$\varepsilon > \frac{\mu}{\sigma}\left[\mu + \sqrt{\mu^2 - \sigma^2}\right].$$

Considérons une variable aléatoire U ne prenant qu'un nombre fini de valeurs $x_1 < x_2 < \ldots < x_n$ avec les probabilités $a_1, a_2, \ldots, a_n$. Si, par exemple, $x_j \leqq \varepsilon \leqq x_{j+1}$, on aura

$$(98) \qquad a_1 + a_2 + \ldots + a_{j-1} \leqq \text{Pr.} \{ U < \varepsilon \} \leqq a_1 + a_2 + \ldots + a_{j+1}.$$

Or la méthode des moments de Tchebicheff étudiée dans ce Traité montre qu'au moyen des formules (1), p. 15 de ce Traité (I, 2), on peut calculer $a_1, a_2, \ldots, a_n$ et $x_1, \ldots, x_n$ connaissant les moments $c_0 = 1, c_1, c_2, \ldots, c_{2n-1}$ de U, et même déterminer ainsi la variable aléatoire U de sorte qu'elle ait les mêmes $2n$ premiers moments $c_0, \ldots, c_{2n-1}$ qu'une variable aléatoire arbitrairement choisie Z. C'est ce qu'on peut faire, par suite, dans le cas où on a pu choisir U de façon à avoir les mêmes $2n-1$ premiers moments que $Z = |X - Y|$, connaissant les $2n-1$ premiers écarts moyens $\lambda^{(r)} = \sqrt[r]{c_r}$, $(r = 1, \ldots, 2n-1)$ de X et de Y.

La méthode de Tchebicheff montre que non seulement on peut faire ce choix de U (I, 2, p. 15), mais encore que la courbe représentative de la fonction des probabilités totales de Z rencontrera la courbe analogue (qui est en escalier) pour U sur chacun de ses morceaux parallèles à l'un des axes (*voir*, par exemple, von Mises, 3, p. 249). Il en résulte qu'on peut remplacer U par Z dans les inégalités (98).

Les inégalités (98) et les expressions des a_j en fonction des $\lambda^{(j)}$ fourniront alors les relations d'inégalités cherchées entre $q(\varepsilon) = \text{Pr.} \{ |X - Y| < \varepsilon \}$ et $\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(2n-1)}$, relations qui seront bien entendu indépendantes de $a_1, \ldots, a_n, x_1, \ldots, x_n$.

Par exemple, prenant $n = 2$, un calcul facile (von Mises, 3, p. 249) fournit les inégalités suivantes dues à Tchebicheff.

$$\text{Pr.} \{ |X - Y| < x_1 \} \leqq a_1, \qquad \text{Pr.} \{ |X - Y| > x_2 \} \geqq a_2$$

avec

$$x_1, x_2 = \frac{\mu}{2}\left[\rho \mp \sqrt{4 + \rho^2}\right], \qquad a_1, a_2 = \frac{1}{2}\left[1 \mp \frac{\rho}{\sqrt{4 + \rho^2}}\right],$$

où $\mu^2 = \mathfrak{M}(X - Y)^2$, $\rho = \left[\frac{\lambda^{(3)}}{u}\right]$.

Revenant à n quelconque, la méthode employée permet aussi de montrer que l'on obtient ainsi, pour les mêmes données $\lambda^{(1)}, \ldots, \lambda^{(2n-1)}$, les formules les plus avantageuses.

La méthode peut d'ailleurs s'étendre, moyennant quelques précautions, au cas où le dernier écart donné est d'ordre pair. On retrouverait ainsi, par exemple, l'inégalité (95).

Moyennes conditionnées. — M. Kolmogoroff a fait un usage intéressant (4, p. 46) de ce qu'on peut appeler les moyennes conditionnées (bedingte Erwartungen), déjà rencontrées (p. 69).

On peut désigner par $\mathfrak{M}_e X$ la valeur moyenne de la variable aléatoire X dans la catégorie de celles des épreuves (sur lesquelles X est défini) où l'événement e a lieu.

Si l'on représente par $C_e(x)$ la probabilité que $X < x$ parmi les épreuves où e a lieu, on voit qu'on aura

$$\mathfrak{M}_e X = \int_{-\infty}^{+\infty} x \, d\, C_e(x).$$

Soit maintenant

$$\mathfrak{M} X = \int_{-\infty}^{+\infty} x \, d\, C(x),$$

d'autre part, supposons que $e_1, e_2, \ldots, e_n$ soient des événements incompatibles et les seuls possibles.

On aura évidemment

$$C(x) = \sum_j [\text{Pr.}\, e_j]\, C_{e_j}(x),$$

d'ou

$$\mathfrak{M} X = \sum_j [\text{Pr}\; e_j]\, \mathfrak{M}_{e_j} X. \tag{99}$$

Application à une inégalité de Kolmogoroff. — Soient $X_1, X_2, \ldots, X_n$, n variables aléatoires *indépendantes* et $Y_k = X_k - \overline{X}_k$; soit enfin μ_k l'écart quadratique moyen de Y_k, et posons $S_k = Y_1 + \ldots + Y_k$. Cherchons une borne supérieure de la probabilité P pour que l'une au moins des inégalités

$$|S_1| \geqq \varepsilon, \quad \ldots, \quad |S_n| \geqq \varepsilon$$

soit vérifiée.

En vertu de l'inégalité (10) de la page 15, on a

$$P \leqq \text{Pr.}\,\{|S_1| \geqq \varepsilon\} + \ldots + \text{Pr}\,\{|S_n| \geqq \varepsilon\}$$

et en vertu de l'inégalité de Bienaymé

$$\Pr\{|S_k| \geqq \varepsilon\} < \frac{\mu_1^2 + \ldots + \mu_k^2}{\varepsilon^2};$$

d'où

$$P \leqq \frac{1}{\varepsilon^2}\{\mu_1^2 + (\mu_1^2 + \mu_2^2) + \ldots + (\mu_1^2 + \ldots + \mu_n^2)\})$$

M. Kolmogoroff (2, p. 310), a obtenu une inégalité plus précise en utilisant la notion de moyenne conditionnée. P est la probabilité pour que se produise l'un des événements incompatibles $e_1, \ldots, e_n$, e_k étant la probabilité pour que l'on ait simultanément

$$|S_1| < \varepsilon, \quad \ldots, \quad |S_{k-1}| < \varepsilon, \quad |S_k| \geqq \varepsilon$$

Or on a, en vertu de (99),

$$\mu_1^2 + \ldots + \mu_n^2 = \mathfrak{M} S_n^2 = \sum_k [\Pr. e_k] \mathfrak{M}_{e_k} S_n^2.$$

Mais

$$\mathfrak{M}_{e_l} S_n^2 = \mathfrak{M}_{e_k}\left[S_k^2 + 2\sum_{j>k} S_k Y_j + \sum_{j>k} Y_j^2 + 2\sum_{j,h>k} Y_j Y_h\right].$$

Comme e_k n'affecte pas $Y_{k+1}, \ldots, Y_n$ et que $Y_{k+1}, \ldots, Y_n$ sont independants de S_k, on a

$$\mathfrak{M}_{e_k} S_k Y_j = [\mathfrak{M}_{e_k} S_k][\mathfrak{M} Y_j] = [\mathfrak{M}_{e_k} S_k] \times 0 = 0 \qquad \text{pour } j > k$$

et de même

$$\mathfrak{M}_{e_l} Y_j Y_h = 0 \qquad \text{pour } \left.\begin{matrix} h \\ j \end{matrix}\right\} > k, \qquad \text{avec } j \neq h$$

Enfin, quand e_k a lieu, $|S_k| \geqq \varepsilon$ et, par suite,

$$\mathfrak{M}_{e_l} S_k^2 \geqq \varepsilon^2, \qquad \text{d'où} \qquad \mathfrak{M}_{e_k} S_n^2 \geqq \varepsilon^2.$$

Finalement,

$$\mu_1^2 + \ldots + \mu_n^2 \geqq \varepsilon^2 \sum_k \Pr. e_k = \varepsilon^2 P.$$

D'où l'*inégalité de Kolmogoroff*

$$P \leqq \frac{1}{\varepsilon^2}(\mu_1^2 + \ldots + \mu_n^2)$$

ou encore : S_k étant la somme $Y_1 + \ldots + Y_k$ de variables *indépen-*

dantes de valeurs moyennes nulles, on a

$$\text{Pr.}\left\{|S_1| \geqq \varepsilon \text{ ou } |S_2| \geqq \varepsilon \text{ ou } \ldots \text{ ou } |S_n| \geqq \varepsilon\right\} \leqq \frac{\mathfrak{M} S_n^2}{\varepsilon^2}, \tag{100}$$

ce qui constitue une généralisation de l'inégalité de Bienaymé, à laquelle elle se réduit pour $n = 1$.

M. Kolmogoroff (5, p. 484) a obtenu aussi d'autres inégalités intéressantes.

D'abord, pour ε positif, fixe, arbitraire,

$$\text{Pr.}\left\{|s_n| > \frac{\mu}{2}\right\} \geqq \frac{1}{1600} \tag{101}$$

en posant $S_n = s_n - \bar{s}_n$ où $s_n = X_1 + \ldots + X_n$ et en appelant μ l'écart quadratique moyen de S_n.

Il est à observer que si l'on retranche une constante de chacune des quantités X_j, le second membre de cette inégalité ne change pas, non plus que μ. L'inégalité sera donc d'autant plus précise qu'on choisira mieux ces constantes. Si, par exemple, s_n a une probabilité élémentaire et si la loi de cette densité de probabilité est symétrique, le minimum de ce premier membre, quand on retranche de s_n une quantité certaine k, aura lieu quand $k = \bar{s}_n$. On aura donc, en général, chance et, au moins dans ce cas particulier, certitude d'obtenir une inégalité plus précise, en remplaçant, dans la dernière égalité, $\text{Pr.}\left\{|s_n| > \frac{\mu}{2}\right\}$ par $\text{Pr.}\left\{|S_n| > \frac{\mu}{2}\right\}$.

Dans le cas où les déviations des variables aléatoires indépendantes X_i sont supposées bornées dans leur ensemble, c'est-à-dire telles que $|X_i - \bar{X}_i| < M$, M étant un nombre indépendant de i et le même d'une épreuve à l'autre, M. Kolmogoroff a établi la formule suivante :

$$\text{Pr.}\left\{|s_n| \geqq \varepsilon\right\} \geqq \frac{1}{1600}\left[1 - \frac{M^2 + 4\varepsilon^2}{\mu^2}\right]. \tag{102}$$

Pour la démonstration de (101), (102) nous nous contenterons de renvoyer à son mémoire.

Une inégalité de Serge Bernstein. — Considérons une somme

$$S = X_1 + \ldots + X_n$$

de n variables aléatoires *indépendantes*, sa valeur moyenne

$$\overline{S} = \overline{X}_1 + \quad + \overline{X}_n$$

et la déviation $\delta = S - \overline{S}$, ainsi que l'écart quadratique moyen μ de δ. D'après l'inégalité de Bienaymé, la probabilité P pour que $|\delta| > t\mu$ est inférieure à $\frac{1}{t^2}$. M. Serge Bernstein a donné une estimation beaucoup plus précise de P, sous la réserve que la distribution des probabilités des X_i satisfasse à une certaine condition (B) et que t ne soit pas trop grand [condition (b)].

M. S. Bernstein (2, p. 155) pose *a priori* la condition

$$\text{(B)} \qquad |\mathfrak{M}Y_j^k| \leqq H^{k-2}\frac{k!}{2}\mathfrak{M}Y_j^2$$

pour une valeur convenable de H et pour tout entier $k \geqq 2$, avec $Y_i = X_i - \overline{X}_i$, et il démontre que si les variables X_i sont indépendantes et vérifient (B), alors

$$\Pr.\{\delta > t\mu\} < e^{-\frac{t^2}{4}}$$

pour

$$(b) \qquad 0 \leqq t \leqq \frac{\mu}{H}.$$

La condition (B) paraît au premier abord un peu mystérieuse ainsi que la condition (b). Il est parfaitement correct, peut-être plus rapide et d'ailleurs suffisant pour des buts pratiques, de poser d'abord ces conditions et de faire la démonstration en les utilisant. Mais c'est peu satisfaisant pour un esprit curieux de comprendre. Nous allons employer une voie inverse où nous rencontrerons ces conditions comme des conditions naturelles et commodes à poser pour arriver au résultat.

Nous appliquerons pour cela l'inégalité de Markoff à la fonction a^δ où a est une constante positive provisoirement arbitraire. On a pour tout $C > 0$

$$\Pr\{a^\delta > C\mathfrak{M}a^\delta\} < \frac{1}{C}$$

ou, en posant $a = e^\alpha$ et supposant $\alpha > 0$, ($a > 1$),

$$\Pr.\left\{\delta > \frac{\mathcal{L}C}{\alpha} + \frac{\mathcal{L}\mathfrak{M}a^\delta}{\alpha}\right\} < \frac{1}{C}.$$

Il était naturel d'essayer d'appliquer cette inégalité à a^δ, parce que

$\mathfrak{M} a^{\delta}$ se calcule facilement :

$$\mathfrak{M} a^{\delta} = \mathfrak{M} a^{Y_1} \mathfrak{M} a^{Y_2} \ldots \mathfrak{M} a^{Y_n},$$

puisque les Y sont indépendants. Il faut, pour la pratique, calculer $\mathfrak{M} a^{Y_j}$ au moyen des moments de Y_j. On écrit donc

$$\begin{aligned} \mathfrak{M} a^{Y_j} = \mathfrak{M} e^{\alpha Y_j} &= \mathfrak{M} \left[1 + \alpha Y_j + \alpha^2 \frac{Y_j^2}{2!} + \ldots \right] \\ &= 1 + \frac{\alpha^2}{2} \mathfrak{M} Y_j^2 + \ldots + \frac{\alpha^k}{k!} \mathfrak{M} Y_j^k + \ldots \\ &= 1 + \frac{\alpha^2}{2} \mathfrak{M} Y_j^2 \left\{ 1 + \ldots + \frac{2\alpha^{k-2}}{k!} \frac{\mathfrak{M} Y_j^k}{\mathfrak{M} Y_j^2} + \ldots \right\}. \end{aligned}$$

Il est alors naturel d'imposer une limitation sur les $\mathfrak{M} Y_j^k$ telle que le calcul de l'accolade soit facile et élimine ces moments supérieurs. On peut, *par exemple*, les supposer tels que l'accolade soit majorée par une progression géométrique convergente. Il faudra pour cela supposer qu'il existe un nombre q tel que

$$\left| \frac{2\alpha^{k-2}}{k!} \frac{\mathfrak{M} Y_j^k}{\mathfrak{M} Y_j^2} \right| \leqq q^{k-2} \qquad \text{avec} \quad q < 1$$

En posant $\frac{q}{\alpha} = \mathrm{H}$, ceci n'est autre que la condition (B). Et en supposant, puisque α est arbitraire, $\alpha \mathrm{H} < 1$, on aura

$$\mathfrak{M} a^{Y_j} \leqq 1 + \frac{\alpha^2}{2(1 - \alpha \mathrm{H})} \mathfrak{M} Y_j^2,$$

d'où

$$\mathcal{L} \mathfrak{M} a^{\delta} = \sum_j \mathcal{L} \left\{ 1 + \frac{\alpha^2}{2(1 - \alpha \mathrm{H})} \mathfrak{M} Y_j^2 \right\} \leqq \sum_j \frac{\alpha^2}{2(1 - \alpha \mathrm{H})} \mathfrak{M} Y_j^2 = \frac{\alpha^2 \mu^2}{2(1 - \alpha \mathrm{H})}.$$

On a donc

$$\Pr. \left\{ \delta > \frac{\mathcal{L} \mathrm{C}}{\alpha} + \frac{\alpha \mu^2}{2(1 - \alpha \mathrm{H})} \right\} \leqq \Pr. \left\{ \delta > \frac{\mathcal{L} \mathrm{C}}{\alpha} + \frac{\mathcal{L} \mathfrak{M} a^{\alpha}}{\alpha} \right\} < \frac{1}{\mathrm{C}}$$

ou, en posant $c = \mathcal{L} \mathrm{C}$,

$$\Pr. \left\{ \delta > \left(\frac{c}{\alpha} + \frac{\alpha \mu^2}{2(1 - \alpha \mathrm{H})} \right) \right\} < e^{-c}.$$

Pour simplifier, prenons α tel que $\alpha \mathrm{H} \leqq \frac{1}{2}$ (qui vérifie bien $\alpha \mathrm{H} < 1$).

On aura $\alpha\mu^2 \geqq \frac{\alpha\mu^2}{2(1-\alpha H)}$; d'où

$$\text{Pr.}\left\{\delta > \left(\frac{c}{\alpha} + \alpha\mu^2\right)\right\} \leqq e^{-c}$$

ou

$$\text{Prob.}\{\delta > \theta\mu\} \leqq e^{-c} \qquad \text{pour} \quad \theta = \frac{c}{\alpha\mu} + \alpha\mu$$

La formule n'a d'intérêt que si $c > 0$. Dans ce cas, la plus petite valeur t, prise par θ pour c fixe, α variable, a lieu pour $\alpha^2\mu^2 = c$: d'où $t = 2\alpha\mu$ et $c = \frac{t^2}{4}$. On a donc, en prenant pour α la valeur $\frac{t}{2\mu}$,

$$\text{Pr.}\{\delta > t\mu\} \leqq e^{-\frac{t^2}{4}} \tag{106}$$

Mais il faudra que $\alpha H \leqq \frac{1}{2}$; d'où la condition ($b$)

$$t \leqq \frac{\mu}{H}.$$

C'est le résultat de Bernstein.

Appliquons-le en remplaçant chaque X_j par $-X_j$, μ ne sera pas changé, la condition (B) sera encore vérifiée, et l'on aura

$$\text{Pr.}\{-\delta > t\mu\} < e^{-\frac{t^2}{4}} \qquad \text{avec} \quad 0 \leqq t < \frac{\mu}{H}. \tag{107}$$

En combinant (106) et (107) on a, sous la condition (B),

$$\text{Pr.}\{|\delta| > t\mu\} < 2e^{-\frac{t^2}{4}} \qquad \text{pour} \quad 0 \leqq t < \frac{\mu}{H}. \tag{108}$$

C'est la formule qu'il convient de comparer avec l'inégalité de Bienaymé. On remarque d'abord que $2e^{-\frac{t^2}{4}}$ n'est < 1 que pour $t > t_0$, t_0 étant un nombre déterminé : $t_0 = 1, 7, \ldots$ De sorte que l'inegalité de Bienaymé

$$\text{Pr.}\{|\delta| > t\mu\} < \frac{1}{t^2}$$

est, des deux formules, la seule utilisable pour $1 < t < t_0 = 1, 7, \ldots$ ou pour $\frac{\mu}{H} < t_0 = 1, 7, \ldots$ La formule de Bernstein est surtout avantageuse pour les grandes valeurs de t. On voit en effet facilement que

$$2e^{-\frac{t^2}{4}} < \frac{1}{t^2} \qquad \text{pour} \quad 0 < t < t_1 = 0, 7, \ldots \quad \text{ou} \quad t > t_2 = 3, \ldots$$

Mais comme elle n'est utile que pour $t > t_0 > t_1$, on voit, en résumé, que, seulement dans le cas où $\frac{\mu}{H} > t_2 = 3. \ldots$, la formule de Bernstein est plus avantageuse que celle de Bienaymé pour $t > t_2 = 3, \ldots$.

Pour montrer la généralité de la condition (B), observons qu'elle est vérifiée quand les $|X_j|$ sont bornés. Plus précisément, supposons qu'il existe un nombre M supérieur ou égal à tous les $|Y_j|$ dans toute épreuve. Représentons $\mathfrak{M} Y_j^k$ sous la forme $\int_{-\infty}^{+\infty} y^k \, dq_j(y)$; on aura

$$|\mathfrak{M} Y_j^k| \leqq \int_{-\infty}^{+\infty} |y^{k-2}| \, y^2 \, dq_j(y) \leqq M^{k-2} \mathfrak{M} Y_j^2.$$

La condition (B) sera donc remplie; car on pourra prendre H tel que

$$M^{k-2} \leqq H^{k-2} \frac{k!}{2} \qquad \text{pour} \quad k \geqq 2$$

Ceci a évidemment lieu pour $H = M$. Toutefois, on a avantage, en vue de rendre moins stricte la condition (b), à prendre H le plus petit possible. Pour $k = 3$, il faut $\frac{M}{H} \leqq 3$. On voit facilement que cela suffit pour k quelconque. On peut donc prendre $H = \frac{M}{3}$, de sorte que la condition (b) devient

$$t < \frac{3\mu}{M}.$$

Pour que la formule de Bernstein devienne utile, il faut que l'on ait $t > t_0$, ce qui exige $\mu > t_0 \frac{M}{3}$, et pour qu'elle soit préférable à celle de Bienaymé, il faut qu'on ait $t > t_2$ et, par suite,

$$\mu > \frac{t_2 M}{3}. \tag{109}$$

La condition (109) n'est certainement pas vérifiée pour $n = 1$, car alors

$$\mu = \sqrt{\mathfrak{M} Y_1^2} \leqq M < \frac{t_2 M}{3}.$$

Mais, comme

$$\mu^2 = \sum_{j=1}^{j=n} \mathfrak{M} Y_j^2,$$

μ^2 ne décroît pas quand n croît et l'on peut espérer que μ finisse par vérifier (109) pour n assez grand.

Toutefois, comme cela n'est pas certain en général, nous donnerons un exemple où cette condition est nécessairement remplie : c'est celui où X_j est la valeur prise à la $j^{\text{ième}}$ épreuve d'une même variable aléatoire bornée X. Alors

$$\mu^2 = n\mathfrak{M}Y^2 = n\mu_0^2,$$

et l'inégalité de Bernstein devient

$$\text{Pr}\left\{|Y_1 + \ldots + Y_n| > t\mu_0\sqrt{n}\right\} < 2e^{-\frac{t^2}{4}} \tag{110}$$

pour

$$0 < t < \frac{3\mu_0\sqrt{n}}{M},$$

μ_0 étant l'écart quadratique moyen de X.

Comme $\varepsilon = t\mu = t\mu_0\sqrt{n}$, la seconde inégalité (condition b) sera certainement satisfaite pour ε donné, quand n est assez grand

L'inégalité (110) sera utile pour n assez grand et t convenable : savoir

$$\sqrt{n} > \frac{Mt_1}{3\mu_0} \qquad \text{et} \qquad t_0 < t$$

elle sera plus avantageuse que celle de Bienaymé pour n assez grand et t convenable . savoir

$$\sqrt{n} > \frac{Mt_2}{3\mu_0} \qquad \text{et} \qquad t_2 < t$$

En particulier, si $X = (1 \text{ ou } 0)$ avec les probabilités respectives p et $q = 1 - p$. S représente la répétition d'un événement E de probabilite constante p, $\overline{S} = np$. $\mu_0 = \sqrt{pq}$, M est au moins égal à la plus grande des deux valeurs $|1-p| = q$ et $|0-p| = p$ et, en prenant pour simplifier $M = 1$, on aura, en appelant $f = \frac{S}{n}$ la fréquence de l'événement E dans n épreuves,

$$\text{Pr.}\left\{|f-p| > t\sqrt{\frac{pq}{n}}\right\} < 2e^{-\frac{t^2}{4}} \qquad \text{pour} \quad 0 < t < 3\sqrt{npq}$$

Dès qu'on aura

$$n > \frac{t_2^2}{9pq} \qquad \left(\text{par exemple } n \geqq 5 \text{ pour } p = q = \frac{1}{2}\right).$$

la formule de Bernstein (110) sera plus avantageuse que celle de Bienaymé pour $t > t_2$.

Faisons encore une vérification; on sait (p. 97) que

$$\text{Pr.}\left\{|f-p| > \lambda\sqrt{\frac{2pq}{n}}\right\}$$

tend vers l'intégrale de Laplace $1-\Theta(\lambda)$. On doit donc avoir, en prenant $t=\lambda\sqrt{2}$,

$$1-\Theta(\lambda) \leqq 2e^{-\frac{\lambda^2}{2}},$$

ce qu'on vérifie facilement en étudiant la dérivée de $1-\Theta(\lambda)-2e^{-\frac{\lambda^2}{2}}$.

D'ailleurs, nous n'avons étudié l'inégalité (110) qu'à titre d'application au cas de Bernoulli de la formule générale de Bernstein (108). Nous avons cité (p. 97) une formule établie directement pour ce cas par M. S. Bernstein lui-même et qui est plus précise que (110).

Formules de Gauss-Winkler et de Camp. — Le raisonnement de la page 119, montre que si l'on veut obtenir une borne supérieure de la probabilité que $|X-Y| > \varepsilon$, où X, Y sont deux variables aléatoires générales, sans en connaître autre chose que leur écart moyen d'un seul ordre, r, déterminé, on ne peut espérer perfectionner la formule d'ordre r (88). Mais, encouragé par le succès de l'hypothèse de Gauss signalée plus haut (p. 62), on pouvait se demander si l'on ne pourrait améliorer cette formule en faisant sur la loi de probabilité envisagée des hypothèses très générales et dont il serait aisé de s'assurer, dans les cas concrets, si elles sont ou non vérifiées. C'est effectivement ce qui a lieu, comme cela a été établi dans l'hypothèse de Gauss par Winkler et divers auteurs et dans un cas plus général par M. Camp (**1**). On peut d'ailleurs (Fréchet, **140**, p. 12), étendre le champ de validité de leurs formules et même rendre celles de M. Camp encore plus précises. Ces raisonnements sont un peu longs, mais élémentaires. Nous n'hésiterons pas à les reproduire ici, car leurs résultats en valent la peine. Sans grande utilité dans la Théorie des probabilités, les formules obtenues peuvent, au contraire, rendre de grands services en Statistique où les lois de probabilité ne sont souvent qu'imparfaitement connues.

Pour abréger, au lieu de reproduire d'abord le raisonnement direct

qui s'applique au cas de Gauss, nous traiterons ce cas comme cas particulier de celui de M. Camp.

Cas de Gauss et de M. Camp. Leurs généralisations. — Il s'agit de déterminer une limite supérieure de la probabilité $\varpi(u)$ pour qu'une variable aléatoire X diffère d'une autre variable aléatoire Y d'au moins u fois l'écart moyen λ_r d'ordre r de X et de Y.

Nous savons déjà que $\varpi(u) \leqq \frac{1}{u^r}$. Pour améliorer ce résultat, plaçons-nous dans le cas où les hypothèses suivantes sont réalisées.

1° Il y a une probabilité élémentaire $f(x)\,dx$ pour que X soit entre x et $x + dx$;

2° La courbe de densité de probabilité $y = f(x)$ a un seul maximum;

3° On prend pour Y l'abscisse certaine, w, de ce maximum.

C'est, à peu près, l'hypothèse de Gauss de la page 62.

Si ces conditions sont remplies, la probabilité pour que $|X - Y| < \varepsilon$, est une fonction de ε,

$$q(\varepsilon) = \int_{w-\varepsilon}^{w+\varepsilon} f(x)\,dx = \int_0^{\varepsilon} [f(w+s) + f(w-s)]\,ds$$

qui est continue, croissante, et dont la représentation graphique est *convexe* (¹) vers le haut. Ces conditions plus générales nous suffiront. Elles présentent sur celles de Gauss le très grand avantage que pour une même variable aléatoire X, et en prenant pour Y une valeur certaine w, la formule peut, dans certains cas, être appliquée à une infinité de valeurs de w et non, comme dans les hypothèses de Gauss, à une seule. Par exemple, si $f(x) = \frac{1}{\pi(1+x^2)}$, ses hypothèses ne sont satisfaites qu'en prenant $Y = w = 0$; or, dans le cas où $w \neq 0$,

$$q(\varepsilon) = \frac{1}{\pi}\int_0^{\varepsilon} \left[\frac{1}{1+(w+t)^2} + \frac{1}{1+(w-t)^2}\right] dt,$$

d'où

$$q''(\varepsilon) = \frac{-4}{\pi}\,\frac{\varepsilon[1 - w^2 + \varepsilon^2]}{[1+(w+\varepsilon)^2]^2\,[1+(w-\varepsilon)^2]^2}.$$

(¹) Comme on le voit, en formant q''_{ε}.

Donc, $q''(\varepsilon) < 0$ pour toutes valeurs > 0 de ε quand $\omega^2 < 1$. Ainsi, la courbe représentative de $q(\varepsilon)$ sera concave vers le bas, non seulement pour $\omega = 0$, mais pour $-1 \leq \omega \leq 1$.

En revenant au cas d'une loi quelconque, observons que si $q(t)$ est représenté par une courbe concave vers le bas, il passe en tout point de cette courbe au moins une droite (la tangente, si elle existe) au-dessous de laquelle se trouve la courbe.

Nous allons établir plus loin les formules mêmes de Gauss-Winkler dans ce que nous appellerons *le cas de Gauss généralisé*. Ce sera le cas où, sans avoir à faire l'hypothèse de l'existence d'une densité de probabilité pour X, ni celle d'un maximum unique de cette densité, nous supposerons simplement ceci :

La courbe représentative de la fonction

$$q(t) = \Pr\{|X - Y| < t\}$$

est convexe vers le haut ou, plus généralement, est bornée vers le haut par au moins une certaine droite passant par le point de cette courbe, dont l'abscisse est le nombre positif ε pour lequel on désire calculer une borne inférieure de $q(\varepsilon)$. On pourrait exprimer cette hypothese en disant que cette courbe est « integralement » convexe vers le haut au point $t = \varepsilon$.

M. Camp s'est proposé d'obtenir une formule présentant sur celles de Gauss-Winkler (il est vrai, aux dépens de sa simplicité analytique) l'avantage de rester valable pour des lois de probabilité plus générales encore.

Supposant comme Gauss qu'il y a encore une probabilité élémentaire $f(x)dx$ que X soit compris entre x et $x + dx$, il suppose que la décroissance de $f(x)$ quand x tend vers $+\infty$ ou $-\infty$, au lieu d'avoir lieu depuis un point unique, a lieu à partir des extrémités d'un intervalle déterminé. Et il prend pour Y l'abscisse ω d'un point de cet intervalle. On peut toujours supposer que ce point en soit le milieu, car, dans le cas contraire, il suffirait de prolonger suffisamment cet intervalle. Nous pouvons donc désigner les extrémités de cet intervalle par $\omega \pm A$. Et la fonction

$$q(\varepsilon) = \Pr.\{|X - Y| < \varepsilon\} = \int_0^\varepsilon [f(\omega + t) + f(\omega - t)]\, dt$$

sera représentée par une courbe concave vers le bas pour $\varepsilon \geq A$.

Il ne nous sera utile pour la suite de retenir, ni l'hypothèse de l'existence d'une densité de la probabilité, ni celle de sa décroissance en s'éloignant de ϖ. Il nous suffira de supposer que la courbe representative de $q(t)$ est concave vers le bas quand t est assez grand $(t \geqq A)$, ou plus généralement qu'elle est convexe vers le haut pour $t \geqq A$, au point $t = \varepsilon$. C'est-à-dire qu'elle est bornée en haut pour t assez grand $(t \geqq A)$ par au moins une droite passant par le point de cette courbe ayant pour abscisse la valeur ε de t pour laquelle on veut calculer une borne inférieure de $q(\varepsilon)$. Si cette unique condition est réalisée, on est dans ce que nous appellerons *le cas de Camp généralisé*.

Formules préparatoires. — On a

$$
\begin{aligned}
(\lambda_r)^r &\geqq \int_A^{+\infty} t^r \, dq(t) \\
&= -\int_A^{+\infty} t^r \, d[1 - q(t)] \\
&= -\{t^r[1 - q(t)]\}_{t=A}^{t=+\infty} + \int_A^{+\infty} p(t) \, dt^r
\end{aligned}
$$

en posant $p(t) = 1 - q(t)$. D'ailleurs, on a

$$\int_B^{+\infty} t^r \, dq(t) \geqq B^r[1 - q(B)]$$

et, puisqu'on postule l'existence de λ_r, le second membre est infiniment petit avec $\frac{1}{B}$. De sorte que, pour tout $B > A$,

$$(\lambda_r)^r \geqq A^r p(A) + \int_A^B p(t) \, dt^r \quad [1]$$

Pour estimer la dernière intégrale, observons que, par hypothèse, il existe (*fig.* 3) une droite $y = m(t - \varepsilon) + p(\varepsilon)$, située au-dessous de

[1] On obtiendrait un résultat encore meilleur en tenant compte de l'intégrale $\int_B^{+\infty} p(t) \, dt^r$ que nous avons supprimée.

la courbe $p = p(t)$ [qui, au contraire de $q = q(t)$, est, en ce point, convexe vers le bas], tout au moins pour $t > A$. Donc

$$(\lambda_r)^r \geqq A^r p(A) + \int_A^B y \, dt^r.$$

Soit maintenant z l'ordonnée, correspondant à l'abscisse t, de la corde passant par les points d'abscisses A et ε. Pour que l'on puisse écrire

$$(111) \qquad (\lambda_r)^r \geqq A^r p(A) + \int_A^B z \, dt^r,$$

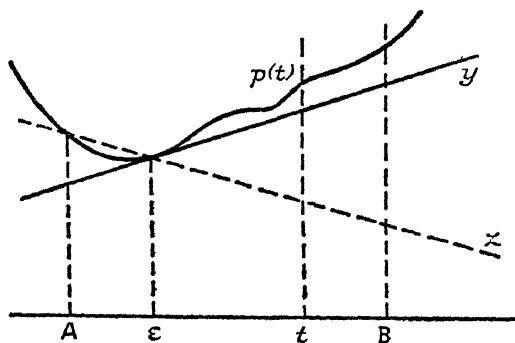



Fig 3.

il suffit de prendre ε, de sorte que

$$\int_A^B (y - z) \, dt^r \geqq 0$$

D'ailleurs, comme $y - z$ est de la forme $m'(t - \varepsilon)$, avec $m' > 0$, il suffit pour cela que

$$0 \leqq \int_A^B (t - \varepsilon) \, dt^r = \frac{r}{r+1}(B^{r+1} - A^{r+1}) - \varepsilon(B^r - A^r)$$

et que $\varepsilon < B$, ou encore que

$$(112) \qquad \varepsilon \leqq \frac{r}{r+1} \frac{B^{r+1} - A^{r+1}}{B^r - A^r} \leqq B.$$

On vérifie facilement que lorsque B croît de A à $+\infty$, le second

membre croît en restant $\leqq B$, de A à $+\infty$. Il passe donc une fois et une seule par la valeur $\varepsilon > A$ pour $B = b$, le nombre b étant défini par la condition

$$\varepsilon = \frac{r}{r+1}\,\frac{b^{r+1}-A^{r+1}}{b^r-A^r} \qquad (b > A). \tag{113}$$

On aura évidemment $\varepsilon \leqq b$.
La condition (112) sera donc satisfaite pour $B \geqq b$.
Nous pouvons maintenant calculer z par

$$\frac{z-p(\varepsilon)}{t-\varepsilon} = \frac{p(\varepsilon)-p(A)}{\varepsilon-A}$$

et, par suite,

$$\int_A^B z\,dt^r = (B^r - A^r)\left\{p(\varepsilon) - \varepsilon\frac{p(\varepsilon)-p(A)}{\varepsilon - A}\right\} + \frac{r}{r+1}(B^{r+1} - A^{r+1})\frac{p(\varepsilon)-p(A)}{\varepsilon - A}.$$

D'où, pour $B \geqq b$,

$$(\lambda_r)^r \geqq A^r p(A) + p(A)(B^r - A^r) + [p(\varepsilon)-p(A)]\left\{(B^r - A^r) + \frac{\left[\frac{r}{r+1}(B^{r+1}-A^{r+1}) - \varepsilon(B^r - A^r)\right]}{\varepsilon - A}\right\}.$$

Pour $B \geqq b \geqq A$, le coefficient de $p(\varepsilon) - p(A)$ est sûrement positif et l'on a

$$p(\varepsilon) \leqq p(A) + \frac{(\lambda_r)^r - p(A)B^r}{B^r - A^r + \dfrac{\frac{r}{r+1}(B^{r+1}-A^{r+1}) - \varepsilon(B^r - A^r)}{\varepsilon - A}}$$

ou

$$p(\varepsilon) \leqq \frac{(\varepsilon - A)(\lambda_r)^r + p(A)\left\{\frac{r}{r+1}(B^{r+1}-A^{r+1}) - A(B^r - A^r) - B^r(\varepsilon - A)\right\}}{\frac{r}{r+1}(B^{r+1}-A^{r+1}) - A(B^r - A^r)}. \tag{114}$$

Dans le second membre, figure $p(A)$ qui est inconnu. On peut l'estimer au moyen de la formule d'ordre r,

$$p(A) \leqq \left(\frac{\lambda_r}{A}\right)^r$$

et alors deux cas se présentent suivant le signe du coefficient

de $p(\mathrm{A})$

$$
(115)\quad \left\{
\begin{array}{l}
\dfrac{r\mathrm{B}^{r+1}+\mathrm{A}^{r+1}}{r+1}-\varepsilon\mathrm{B}^r \leqq 0 \qquad p(\varepsilon)\leqq \dfrac{(\varepsilon-\mathrm{A})(\lambda_r)^r}{\dfrac{r}{r+1}(\mathrm{B}^{r+1}-\mathrm{A}^{r+1})-\mathrm{A}(\mathrm{B}^r-\mathrm{A}^r)}, \\[2ex]
\dfrac{r\mathrm{B}^{r+1}+\mathrm{B}^{r+1}}{r+1}-\varepsilon\mathrm{B}^r > 0\cdot \\[2ex]
\qquad p(\varepsilon)\leqq\left(\dfrac{\lambda_r}{\mathrm{A}}\right)^r\left\{\dfrac{\dfrac{r}{r+1}(\mathrm{B}^{r+1}-\mathrm{A}^{r+1})-\varepsilon(\mathrm{B}^r-\mathrm{A}^r)}{\dfrac{r}{r+1}(\mathrm{B}^{r+1}-\mathrm{A}^{r+1})-\mathrm{A}(\mathrm{B}^r-\mathrm{A}^r)}\right\}.
\end{array}
\right.
$$

D'ailleurs, dans le second cas, si $\mathrm{A}\leqq\lambda_r$, on améliorerait encore la formule en remplaçant $p(\mathrm{A})$ par l'unité.

Nous venons de mener le calcul de façon à préparer les généralisations annoncées. Mais appliquons-le d'abord comme M. Camp

Calcul de M. Camp. — M. Camp a obtenu sa formule en mettant en évidence un terme ω à calculer une fois pour toutes et en prenant pour B une valeur particulière. Représentons le coefficient de $p(\mathrm{A})$ dans (114) par $\frac{\omega}{1+\omega}$, c'est-à-dire posons

$$
(116)\qquad \omega=\frac{\dfrac{r\mathrm{B}^{r+1}+\mathrm{A}^{r+1}}{r+1}-\varepsilon\mathrm{B}^r}{\mathrm{B}^r(\varepsilon-\mathrm{A})}=\frac{\dfrac{\mathrm{A}^{r+1}}{r+1}+\mathrm{B}^r\left(\dfrac{r}{r+1}\mathrm{B}-\varepsilon\right)}{\mathrm{B}^r(\varepsilon-\mathrm{A})}
$$

Alors (114) devient

$$
(117)\qquad p(\varepsilon)\leqq\frac{1}{1+\omega}\left(\frac{\lambda_r}{\mathrm{B}}\right)^r+p(\mathrm{A})\frac{\omega}{1+\omega}\qquad \text{avec } 1+\omega>0,
$$

de sorte que

$$
(118)\qquad p(\varepsilon)\leqq\frac{1}{1+\omega}\left(\frac{\lambda_r}{\mathrm{B}}\right)^r\qquad \text{si } \omega\leqq 0,
$$

$$
(118^{bis})\qquad p(\varepsilon)\leqq\frac{\left(\dfrac{\mathrm{A}}{\mathrm{B}}\right)^r+\omega}{1+\omega}\left(\frac{\lambda_r}{\mathrm{A}}\right)^r\qquad \text{si } \omega\geqq 0.
$$

Pour $\mathrm{A}\leqq\lambda_r$ (et $\omega\geqq 0$), on peut même améliorer, en remplaçant dans (117), $p(\mathrm{A})$ par 1 et non par $\left(\frac{\lambda_r}{\mathrm{A}}\right)^r$. D'où

$$
p(\varepsilon)\leqq\frac{\left(\dfrac{\lambda_r}{\mathrm{B}}\right)^r+\omega}{1+\omega}.
$$

On observe que l'expression de ω se simplifierait si l'on pouvait choisir pour B [uniquement assujetti à la condition $B \geqq b$, ou encore à (112)] la valeur $B = \frac{r+1}{r}\varepsilon$ Or, il est facile de voir qu'en prenant $\varepsilon = \frac{rB}{r+1}$, l'inégalité (112) est satisfaite.

Pour ce choix de B, qui est le choix de M. Camp, on a

$$\omega = \frac{1}{\varepsilon - A}\left(\frac{r}{\varepsilon}\right)^r \left(\frac{A}{r+1}\right)^{r+1} > 0$$

et, si l'on pose $h = \frac{A}{\varepsilon} < 1$,

$$\omega = \frac{r^r}{(r+1)^{r+1}} \frac{h^{r+1}}{(1-h)} > 0.$$

D'où *la formule de Camp*

$$p(\varepsilon) \leqq \left(\frac{r}{r+1}\frac{\lambda_r}{\varepsilon}\right)^r (1+\eta) \qquad (118^{ter})$$

avec

$$\eta = \frac{h\left[\left(\frac{r+1}{r}\right)^r - h^r\right]}{(1-h)\frac{(r+1)^{r+1}}{r^r} + h^{r+1}}.$$

Pour chaque valeur de r, on peut dresser, avec M. Camp, un tableau des valeurs de η suivant les valeurs de $h = \frac{A}{\varepsilon}$.

On aura, par exemple, pour $r = 1$,

$$\eta = \frac{h(2-h)}{4(1-h)+h^2} = \frac{h}{2-h},$$

$$p(\varepsilon) \leqq \frac{1}{2}\frac{\sigma}{\varepsilon}(1+\eta) = \frac{\sigma}{\varepsilon}\frac{1}{2-h}.$$

h .	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$
η	1	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{9}$	$\frac{1}{11}$
$\frac{1+\eta}{2}$	1	$\frac{2}{3}$	$\frac{3}{5}$	$\frac{4}{7}$	$\frac{5}{9}$	$\frac{6}{11}$
$p(\varepsilon) \leqq$	$\frac{\sigma}{\varepsilon}$	$\frac{2}{3}\frac{\sigma}{\varepsilon}$	$\frac{3}{5}\frac{\sigma}{\varepsilon}$	$\frac{4}{7}\frac{\sigma}{\varepsilon}$	$\frac{5}{9}\frac{\sigma}{\varepsilon}$	$\frac{6}{11}\frac{\sigma}{\varepsilon}$

Pour $r = 2$, la décroissance de η avec h serait plus rapide.

Cas de Gauss généralisé. — Appliquons d'abord la formule générale (114) dans le cas où $A = 0$, c'est-à-dire dans le cas de Gauss généralisé.

Quand $A = 0$, cette formule devient, puisque $p(A) = p(0) = 1$,

$$p(\varepsilon) \leqq \frac{\frac{r+1}{r}\varepsilon(\lambda_r)^r}{B^{r+1}} + 1 - \frac{\varepsilon\left(\frac{r+1}{r}\right)}{B} \qquad \text{pour } B \geqq b$$

Or, l'équation (113) qui fournit b devient ici

$$b = \frac{r+1}{r}\varepsilon;$$

d'où

$$p(\varepsilon) \leqq \frac{b(\lambda_r)^r}{B^{r+1}} + 1 - \frac{b}{B} \qquad \text{pour } B \geqq b.$$

Le second membre est une fonction $U(B)$ de B qui, lorsque B croît à partir de zéro, décroît d'abord, passe par le minimum

$$(119) \qquad 1 - \frac{\varepsilon}{(r+1)^{\frac{1}{r}}\lambda_r} \qquad \text{pour } B = (r+1)^{\frac{1}{r}}\lambda_r,$$

puis croît constamment. On a donc à distinguer deux cas. Tout d'abord si

$$(r+1)^{\frac{1}{r}}\lambda_r \leqq b,$$

c'est-à-dire si

$$\varepsilon \geqq \frac{r}{(r+1)^{\frac{r-1}{r}}}\lambda_r,$$

la plus petite valeur prise par $U(B)$ pour $B \geqq b$ est

$$U(b) = \left(\frac{\lambda_r}{b}\right)^r = \left(\frac{r}{r+1}\,\frac{\lambda_r}{\varepsilon}\right)^r,$$

et l'on a *la première formule*

$$(120) \qquad p(\varepsilon) \leqq \left(\frac{r}{r+1}\,\frac{\lambda_r}{\varepsilon}\right)^r$$

ou, si $\varpi(u)$ est la probabilité que $|X - \mathfrak{w}| < u\lambda$,

$$\varpi(u) \leqq \frac{1}{\left[\left(1 + \frac{1}{r}\right)u\right]^r}.$$

D'ailleurs, ces deux formules ne sont utiles que si elles donnent une borne inférieure à l'unité, c'est-à-dire si

$$\varepsilon > \frac{r}{r+1}\lambda_r \qquad \text{ou} \qquad u > \frac{r}{r+1}.$$

Supposons maintenant

$$\varepsilon < \frac{r}{(r+1)^{1-\frac{1}{r}}}\lambda_r,$$

alors la plus petite valeur de U(B) pour $B \geqq b$ est la valeur (119), et l'on a

$$\varpi(u) \leqq 1 - \frac{u}{(r+1)^{\frac{1}{r}}},$$

$$(121) \qquad p(\varepsilon) \leqq 1 - \frac{\varepsilon}{(r+1)^{\frac{1}{r}}\lambda_r}$$

borne de $p(\varepsilon)$ qui sera, d'après ce qui précède, nécessairement plus petite que celle fournie par la formule (120).

Ainsi, la formule (120) est valable pour toute valeur positive de ε. Mais dans le cas 2°, par l'application de la méthode même qui nous a fourni la formule (120) nous obtenons une formule plus avantageuse (p. 147).

En résumé, nous avons [1].

$$(122) \qquad \boxed{p(\varepsilon) \leqq 1 - \frac{\varepsilon}{(r+1)^{\frac{1}{r}}\lambda_r}, \qquad \text{si } \varepsilon \leqq \frac{r}{(r+1)^{1-\frac{1}{r}}}\lambda_r.}$$

$$(123) \qquad \boxed{p(\varepsilon) \leqq \left(\frac{r}{r+1}\,\frac{\lambda_r}{\varepsilon}\right)^r, \qquad \text{si } \frac{r}{(r+1)^{1-\frac{1}{r}}}\lambda_r \leqq \varepsilon.}$$

La formule (122) présente le grand avantage de fournir une borne supérieure de $p(\varepsilon)$ qui est plus petite que l'unité, *même pour de petites valeurs de* ε, ce qui n'a lieu ni pour la formule de Bienaymé ni pour (123).

---

[1] Les formules (125) relatives au cas de $r = 2$ ont été données sans démonstration

On a, en particulier, pour le cas de $r=1$, $\lambda_1=\sigma$ et $r=2$, $\lambda_2=\mu$,

(124)
$$p(\varepsilon)\leqq 1-\frac{\varepsilon}{2\sigma}, \quad \text{si } \varepsilon\leqq\sigma,$$
$$p(\varepsilon)\leqq \frac{\sigma}{2\varepsilon}, \quad \text{si } \varepsilon\geqq\sigma$$

(125)
$$p(\varepsilon)\leqq 1-\frac{\varepsilon}{\sqrt{3}\,\mu}, \quad \text{si } \varepsilon\leqq\frac{2\mu}{\sqrt{3}};$$
$$p(\varepsilon)\leqq \frac{4}{9}\frac{\mu^2}{\varepsilon^2}, \quad \text{si } \varepsilon\geqq\frac{2\mu}{\sqrt{3}}.$$

A titre d'exemple :

Quand ε atteint en décroissant la valeur μ, la formule de Bienaymé donnant $p(\varepsilon)\leqq 1$ devient illusoire, et la seconde formule (125) donne $p(\varepsilon)\leqq\frac{4}{9}$, ce qui est une notable limitation. Quand, décroissant encore, ε atteint la valeur $\frac{2\mu}{3}$, la seconde formule (125) devient également illusoire, et la première formule (125), soit $p(\varepsilon)\leqq 1-\frac{\varepsilon}{\mu\sqrt{3}}$ donne $p(\varepsilon)\leqq 1-\frac{2}{3\sqrt{3}}$ ou $p(\varepsilon)\leqq 0,615$.

En prenant pour exemple le cas où la loi des écarts de $|x-\alpha|$ obéirait à la loi de probabilité de Laplace, c'est-à-dire où

$$q(t)=\frac{\sqrt{2}}{\mu\sqrt{\pi}}\int_0^t e^{-\frac{x^2}{2\mu^2}}\,dx,$$

on trouverait

$$p\left(\frac{2\mu}{3}\right)=0,544, \qquad p\left(\frac{5\mu}{4}\right)=0,212,$$

---

par Gauss (1, Art. 10) et les formules plus générales (122), (123) par Winkler (1) La première démonstration rigoureuse en a été donnée en 1922 simultanément par M Birger Meidell (1) pour l'inégalité (123) et par M. Faber (1, p. 7), pour (123) et (122), tous les deux dans le cas de Gauss, puis par M. von Mises (3, p. 69) en 1931 dans le cas de Gauss généralisé. Nous avions également donné une démonstration de ces formules dans le cas de Gauss généralisé au cours de nos leçons à l'Institut Henri Poincaré, pendant l'hiver de 1928 et précisé cette extension dans un mémoire de 1931 (Fréchet, 140). *Voir* aussi S Narumi (1).

dont les valeurs ne sont pas très éloignées des bornes supérieures

$$1-\frac{2}{3\sqrt{3}}=0,615 \qquad \text{et} \qquad \frac{4}{9}\times\left(\frac{4}{5}\right)^2=0,284$$

fournies respectivement par les deux dernières formules.

Il y a d'ailleurs lieu d'observer que la formule

$$p(\varepsilon)\leqq 1-\frac{\varepsilon}{(r+1)^{\frac{1}{r}}\lambda_r} \qquad \text{pour } \varepsilon\leqq\frac{r}{(r+1)^{1-\frac{1}{r}}}\lambda_r$$

ne peut être améliorée dans les circonstances où nous nous sommes placés.

Supposons qu'il puisse exister une formule plus précise, dans les cas où la précédente est valable, entendant par là qu'il existerait une fonction $G(\varepsilon, \lambda_r)$ indépendante du couple X, Y et telle qu'on ait

$$(126) \qquad p(\varepsilon)\leqq G(\varepsilon, \lambda_r)\leqq 1-\frac{\varepsilon}{(r+1)^{\frac{1}{r}}\lambda_r} \qquad \text{pour } \varepsilon\leqq\frac{r}{(r+1)^{1-\frac{1}{r}}}\lambda_r$$

quand la loi de probabilité de $|X-Y|$ rentre dans le cas de Gauss généralisé. Alors ces inégalités auraient lieu si cette loi était telle qu'il y eût une densité de probabilité $\delta(t)$ constante et égale à c pour $-a\leqq t\leqq a$ et à zero pour $|t|>a$. De sorte que

$$q(\varepsilon)=2\int_0^{\varepsilon}\delta(t)\,dt=\begin{cases}2c\varepsilon & \text{pour } 0\leqq\varepsilon\leqq a,\\ 2ca & \text{pour } \quad \varepsilon\geqq a\end{cases}$$

La courbe $y=q(x)$ sera bien convexe vers le haut.

Pour avoir $q(+\infty)=1$, il faudra d'ailleurs prendre $2ca=1$.

Alors

$$(\lambda_r)^r=2\int_0^a t^r\,\delta(t)\,dt=2c\frac{a^{r+1}}{r+1}=\frac{a^r}{r+1}$$

et

$$p(\varepsilon)=1-q(\varepsilon)=1-\frac{\varepsilon}{a}=1-\frac{\varepsilon}{(r+1)^{\frac{1}{r}}\lambda_r} \qquad \text{pour } \varepsilon\leqq a;$$

donc, *a fortiori*, pour

$$\varepsilon\leqq\frac{r\lambda_r}{(r+1)^{1-\frac{1}{r}}}=\frac{ar}{r+1}.$$

Les inégalités admises montrent maintenant que

$$G(\varepsilon, \lambda_r) \equiv 1 - \frac{\varepsilon}{(r+1)^{\frac{1}{r}} \lambda_r} \cdot$$

Cas de Camp généralisé. — La formule de Camp, comme la formule (123) ont été obtenues dans ce qui précède en prenant pour B la valeur particulière $\frac{r+1}{r}\varepsilon$ dans la formule (114). Il est naturel de chercher à choisir B de façon à tirer tout le profit possible de celle-ci dans le cas de $A \neq 0$ comme dans le cas de $A = 0$ que nous venons de traiter.

Nous avons alors à considérer deux cas, suivant que le coefficient de $p(A)$ est négatif ou positif.

I. Dans le premier cas, on a, d'après (115).

$$p(\varepsilon) \leqq \frac{(\varepsilon - A)(\lambda_r)^r}{B^r\left[\frac{r}{r+1} B - A\right] + \frac{A^{r+1}}{r+1}} \qquad \text{pour } \varepsilon \geqq A \tag{127}$$

et

$$\frac{A^{r+1}}{B^r(r+1)} + \frac{r}{r+1} B \leqq \varepsilon \leqq \frac{r}{r+1} \frac{B^{r+1} - A^{r+1}}{B^r - A^r}.$$

Observons d'abord que l'inégalité $B \geqq A > 0$, entraîne

$$A \leqq \frac{A^{r+1}}{B^r(r+1)} + \frac{r}{r+1} B \leqq \frac{r}{r+1} \frac{B^{r+1} - A^{r+1}}{B^r - A^r}$$

comme on s'en assure facilement, de sorte que les conditions précédentes sont compatibles.

On a vu (p. 141) que l'inégalité

$$\varepsilon \leqq \frac{r}{r+1} \frac{B^{r+1} - A^{r+1}}{B^r - A^r} \tag{128}$$

est équivalente à $B \geqq b$ (avec $b \geqq A$, quand $\varepsilon \geqq A$). De même la fonction de B

$$\frac{A^{r+1}}{(r+1)B^r} + \frac{r}{r+1} B$$

croissant de A à $+\infty$ en même temps que B passe une fois et une seule par la valeur $\varepsilon \geqq A$ pour une valeur $c > A$ de B (il est clair que

$b \leqq c$). Dès lors l'inégalité

$$\frac{A^{r+1}}{(r+1)B^r} + \frac{r}{r+1} B \leqq \varepsilon$$

se traduit par l'inégalité $B \leqq c$ où c est l'unique racine ($\geqq A$) de l'équation en B

$$(129) \qquad \varepsilon = \frac{A^{r+1}}{(r+1)B^r} + \frac{r}{r+1} B.$$

En résumé, nous nous sommes actuellement placés dans le cas où

$$b \leqq B \leqq c.$$

Pour tirer de l'inégalité (127), valable dans ce cas, le parti le plus avantageux, il faut remplacer le second membre par la valeur minimum qu'il atteint quand B varie entre b et c. Or son dénominateur

$$T(B) = B^r \left[\frac{r}{r+1} B - A\right] + \frac{A^{r+1}}{r+1}$$

est une fonction de B qui croît constamment par valeurs positives quand B croît depuis A. Par suite, il a quand $A \leqq b \leqq B \leqq c$ son maximum pour $B = c$. On déduit donc de (127)

$$p(\varepsilon) \leqq \frac{(\varepsilon - A)(\lambda_r)^r}{T(c)}.$$

D'ailleurs on vérifie qu'en vertu de l'équation de définition de c,

$$T(c) = c^r(\varepsilon - A)$$

Finalement, on a tiré de (127) pour $b \leqq B \leqq c$

$$(130) \qquad p(\varepsilon) \leqq \left(\frac{\lambda_r}{c}\right)^r.$$

II. Plaçons-nous maintenant dans le second cas, celui où le coefficient de $p(A)$ étant $\geqq 0$, on a, d'après (115),

$$(131) \qquad p(\varepsilon) \leqq \left(\frac{\lambda_r}{A}\right)^r \frac{\frac{r}{r+1}(B^{r+1} - A^{r+1}) - \varepsilon(B^r - A^r)}{\frac{r}{r+1}(B^{r+1} - A^{r+1}) - A(B^r - A^r)}$$

pour

$$A \leqq \varepsilon \leqq \frac{A^{r+1}}{B^r(r+1)} + \frac{r}{r+1} B \leqq \frac{r}{r+1} \frac{B^{r+1} - A^{r+1}}{B^r - A^r},$$

c'est-à-dire pour

$$\varepsilon \geqq A \quad \text{et} \quad B \geqq c \geqq b$$

On a vu (p. 142) que si $A \leqq \lambda_r$, on peut obtenir une meilleure inégalité pour $p(\varepsilon)$. Mais la présente reste valable dans ce cas. Pour utiliser au maximum cette inégalité (131), cherchons le minimum du second membre. Celui-ci peut s'écrire

$$L = \left(\frac{\lambda_r}{A}\right)^r \frac{\rho - \varepsilon}{\rho - A}$$

avec

$$\rho = \frac{r}{r+1} \frac{B^{r+1} - A^{r+1}}{B^r - A^r}.$$

Or comme nous l'avons déjà observé page 141, ρ croît de A à $+\infty$ quand B croît de A à $+\infty$ et ρ passe par la valeur ε pour $B = b$ Lorsque B croît à partir de b, L croît donc à partir de zéro et quand B reste $\geqq c$, le minimum de L est atteint pour $B = c$.

Or en tenant compte de l'équation de définition de c on trouve facilement que, pour $B = c$,

$$\rho - A = \frac{A(\varepsilon - A)}{A + rc - (r+1)\varepsilon}$$

et comme

$$\frac{\rho - \varepsilon}{\rho - A} = 1 - \frac{\varepsilon - A}{\rho - A},$$

on voit que, pour $B = c$,

$$L = \left(\frac{\lambda_r}{A}\right)^r \left[1 - \frac{A + rc - (r+1)\varepsilon}{A}\right] = \left(\frac{\lambda_r}{A}\right)^r \frac{(r+1)\varepsilon - rc}{A} = \left(\frac{\lambda_r}{c}\right)^r.$$

Ainsi de l'inégalité (131), pour $B \geqq c$, on tire encore

$$p(\varepsilon) \leqq \left(\frac{\lambda_r}{c}\right)^r. \tag{130}$$

Cette inégalité (130) est valable pour toute valeur de $A \geqq 0$ $\left[\text{dans le cas de } A = 0, \text{ on a } c = (r+1)\frac{\varepsilon}{r} \text{ et l'on retombe sur l'inégalité (123)}\right]$. Mais si $A \leqq \lambda_r$, nous avons vu que l'on pouvait améliorer l'inégalité (114) relative au cas où $B \geqq c$, en remplaçant dans (114), $p(A)$ par l'unité. Examinons ce cas. On a

$$p(\varepsilon) \leqq \frac{(\varepsilon - A)(\lambda_r)^r + \dfrac{r}{r+1}(B^{r+1} - A^{r+1}) - \varepsilon(B^r - A^r) - A^r(\varepsilon - A)}{\dfrac{r}{r+1}(B^{r+1} - A^{r+1}) - A(B^r - A^r)} = M(B)$$

avec

$$A \leqq \lambda_r, \qquad A \leqq \varepsilon, \qquad B \geqq c > b.$$

Or

$$M(B) = 1 - (\varepsilon - A) \left\{ \frac{B^r - (\lambda_r)^r}{\frac{r}{r+1}(B^{r+1} - A^{r+1}) - A(B^r - A^r)} \right\} = 1 - \frac{\varepsilon - A}{N(B)}$$

avec

$$N(B) = \frac{\frac{r}{r+1}(B^{r+1} - A^{r+1}) - A(B^r - A^r)}{B^r - (\lambda_r)^r}.$$

Il s'agit de trouver le minimum de $M(B)$ quand $B \geqq c$ Il a lieu avec celui de $N(B)$. Or,

$$N(B) < 0, \qquad M(B) > 1 \qquad \text{pour } B < \lambda_r,$$
$$M(B) \leqq 1 \qquad \text{pour } B \geqq \lambda_r$$

Il suffit donc de chercher le minimum de $N(B)$ pour les valeurs de B supérieures à la fois à c et à λ_r. Or la dérivée N'_B de $N(B)$ est du signe de

$$\frac{B^{r+1} - A^{r+1}}{r+1} - (\lambda_r)^r (B - A)$$

qui décroît à partir de zéro quand B croît de λ à λ_r et croît ensuite d'une valeur négative à $+\infty$ quand B croît de λ_r à $+\infty$. Par suite, il y a un seul nombre $\alpha > \lambda_r > A$ vérifiant l'équation

$$\frac{B^{r+1} - A^{r+1}}{r+1} - (\lambda_r)^r (B - A) = 0$$

Et $N(B)$ décroît, d'abord, quand B croît de λ_r à α, puis croît ensuite quand B croît depuis α Seulement, nous devons nous limiter aux valeurs de $B \geqq c$, de sorte que nous avons deux cas à examiner :

1° $\alpha \leqq c$. — Alors quand $B \geqq c \geqq \alpha > \lambda_r > A$, le minimum de $N(B)$ et par suite, celui de $M(B)$, sont atteints pour $B = c$. Or cette valeur est celle qui annulait le coefficient de $p(A)$ dans (114) et nous retombons sur la formule déjà obtenue

$$p(\varepsilon) \leqq \left(\frac{\lambda_r}{c}\right)^r.$$

2° $\alpha > c$. — Comme on a aussi $\alpha > \lambda_r$, on voit que quand B varie de façon à rester supérieur à c et à λ_r, le minimum de $M(B)$ est

atteint pour $B = \alpha$. On a donc dans ce cas

$$p(\varepsilon) \leqq M(\alpha) \quad \text{avec } M(\alpha) < M(c)$$

De sorte que la formule obtenue est plus avantageuse que la précédente. On vérifie d'ailleurs facilement que $N(\alpha) = \alpha - A$, de sorte que

$$M(\alpha) = 1 - \frac{\varepsilon - A}{\alpha - A}.$$

D'où

$$p(\varepsilon) \leqq 1 - \frac{\varepsilon - A}{\alpha - A}$$

ou

$$(132) \qquad p(\varepsilon) \leqq \frac{\alpha - \varepsilon}{\alpha - A}.$$

D'ailleurs, c est l'unique racine supérieure à A de l'équation en B,

$$\varepsilon = \frac{A^{r+1}}{(r+1)B^r} + \frac{r}{r+1} B,$$

dans laquelle le second membre est une fonction croissante de B pour $B > A$. La condition $\alpha > c$ se traduit donc par

$$(133) \qquad \varepsilon < \frac{A^{r+1}}{(r+1)\alpha^r} + \frac{r}{r+1} \alpha,$$

où α est l'unique racine supérieure à A (si $A < \lambda_r$) de l'équation

$$(134) \qquad \frac{\alpha^{r+1} - A^{r+1}}{r+1} - (\lambda_r)^2(\alpha - A) = 0 \quad (^1).$$

$\left(\text{Observons que } \frac{A^{r-1}}{(r+1)\alpha^r} + \frac{r}{r+1} \alpha < \alpha, \text{ de sorte qu'on aura } \varepsilon < \alpha\right)$.

Lorsqu'on a calculé c avant α, on peut aussi substituer, à la condition $\alpha > c$, la condition équivalente suivante qui permettra d'éviter le calcul de α

$$(\lambda_r)^r (c - A) - \frac{c^{r+1} - A^{r+1}}{r+1} \geqq 0.$$

---

(1) La condition (133) conjuguée avec la relation (134) se traduit évidemment par une condition de la forme

$$\varepsilon < \varphi(A, \lambda_r, r).$$

Nous expliciterons celle-ci, plus loin, pour $r = 1$. Mais déjà pour $r = 2$, on pourra s'assurer que la fonction φ est chargée de radicaux; on ne peut espérer la mettre sous forme simple pour une valeur positive arbitraire de r.

Les nouvelles formules. — En résumé, si l'on se place dans le cas de Camp généralisé (défini page 139), où $A \geqq 0$ et $\varepsilon \geqq A$, et si l'on appelle c l'unique racine supérieure à A de l'équation

$$\frac{A^{r+1}}{c^r} + rc = (r+1)\varepsilon,$$

on a toujours

$$(130) \qquad \boxed{p(\varepsilon) \leqq \left(\frac{\lambda_r}{c}\right)^r.}$$

Si, en outre, $A < \lambda_r$ et si, en désignant par α l'unique racine supérieure à A de l'équation.

$$(135) \qquad \alpha^{r+1} - A^{r+1} - (r+1)(\lambda_r)^r(\alpha - A) = 0,$$

on a

$$\varepsilon < \frac{A^{r+1}}{(r+1)\alpha^r} + \frac{r}{r+1}\alpha,$$

alors, on a

$$(136) \qquad \boxed{p(\varepsilon) \leqq \frac{\alpha - \varepsilon}{\alpha - A},}$$

et cette inégalité est plus avantageuse que la précédente

C'est d'ailleurs *la plus avantageuse qu'on puisse formuler dans les conditions où nous nous sommes placés*. En d'autres termes, s'il existe une fonction $G(A, \lambda_r, \varepsilon)$ telle que

$$p(\varepsilon) \leqq G(A, \lambda_r, \varepsilon) \leqq \frac{\alpha - \varepsilon}{\alpha - A} \qquad \text{pour } \lambda_r > A \text{ et } \varepsilon > A,$$

où α est l'unique racine superieure à A. de (135), pour tout couple X, Y dont la loi de probabilité est dans le cas de Camp généralisé, alors on doit nécessairement avoir

$$G(A, \lambda_r, \varepsilon) = \frac{\alpha - \varepsilon}{\alpha - A}$$

quand

$$\varepsilon < \left(\frac{A^{r+1}}{\alpha^r} + r\alpha\right)\frac{1}{r+1}.$$

En effet, prenons X, Y tels que

$$q(t) = 0 \qquad \text{pour } 0 \leqq t \leqq A,$$
$$q(t) = N(t - A) \qquad \text{quand } A \leqq t \leqq A'$$

et

$$q(t) = N(A' - A) = 1 \qquad \text{quand } t \geqq A'.$$

La courbe $y = q(x)$ est convexe vers le haut, non pas intégralement (cas de Gauss généralisé) mais pour $x < A$ (cas de Camp généralisé).

Alors

$$(\lambda_r)' = \int_A^{A'} t^r N\, dt = N\frac{A'^{r+1} - A^{r+1}}{r+1};$$

d'où

$$(\lambda_r)' = \frac{1}{r+1}\,\frac{A'^{r+1} - A^{r+1}}{A' - A}$$

et, par suite,

$$\alpha = A'.$$

Or

$$q(\varepsilon) = \frac{\varepsilon - A}{A' - A} \qquad \text{pour } A \leqq \varepsilon \leqq A';$$

d'où

$$p(\varepsilon) = \frac{\alpha - \varepsilon}{\alpha - A}.$$

D'ailleurs,

$$\left(\frac{A^{r+1}}{\alpha^r} + r\alpha\right)\frac{1}{r+1} < \alpha$$

Donc si

$$A \leqq \varepsilon \leqq \left(\frac{A^{r+1}}{\alpha^r} + r\alpha\right)\frac{1}{r+1},$$

on a aussi

$$A \leqq \varepsilon \leqq A'$$

et, par suite,

$$\frac{\alpha - \varepsilon}{\alpha - A} = p(\varepsilon) \leqq G(A, \lambda_r, \varepsilon) \leqq \frac{\alpha - \varepsilon}{\alpha - A};$$

d'où

$$G(A, \lambda_r, \varepsilon) = \frac{\alpha - \varepsilon}{\alpha - A}.$$

Cas particuliers. — 1° On vérifie aisément que, pour $A = 0$, on retombe sur les résultats déjà obtenus page 146.

2° $r = 1$, $\lambda_1 = \sigma$. — On a toujours

$$p(\varepsilon) \leqq \frac{\sigma}{c} \tag{137}$$

en appelant c l'unique racine supérieure à A de l'équation

$$\frac{A^2}{c} + c = 2\varepsilon, \tag{138}$$

soit

$$c = \varepsilon + \sqrt{\varepsilon^2 - A^2};$$

d'où

$$(137^{bis}) \qquad \boxed{p(\varepsilon) \leqq \frac{\sigma}{\varepsilon + \sqrt{\varepsilon^2 - A^2}}.}$$

Si, en outre, $A < \sigma$ et si

$$A < \varepsilon < \frac{1}{2}\left[\frac{A^2}{2\sigma - A} + 2\sigma - A\right],$$

alors on a

$$(139) \qquad \boxed{p(\varepsilon) \leqq 1 - \frac{\varepsilon - A}{2(\sigma - A)}} \qquad \left(< \frac{\sigma}{c}\right).$$

3° $r = 2$, $\lambda_2 = \mu$. — On a toujours

$$(140) \qquad p(\varepsilon) \leqq \frac{\mu^2}{c^2}$$

en appelant c l'unique racine supérieure à A de l'équation

$$(141) \qquad \frac{A^3}{c^2} + 2c = 3\varepsilon$$

Si, en outre, $A < \mu$ et si

$$\varepsilon < \frac{1}{3}\left(\frac{A^3}{\alpha^2} + 2\alpha\right)$$

en désignant par α l'unique racine supérieure à A de l'équation

$$(142) \qquad \alpha^2 + \alpha A + A^2 = 3\mu^2,$$

alors

$$(143) \qquad p(\varepsilon) \geqq \frac{\alpha - \varepsilon}{\alpha - A} = 1 - \frac{2(\varepsilon - A)}{\sqrt{3(4\mu^2 - A^2)} - 3A} \qquad \left(< \frac{\mu^2}{c^2}\right)$$

Calcul numérique. — En revenant au cas d'une valeur quelconque de r, on peut diminuer l'inconvénient qui résulte au point de vue du calcul numérique, de la nécessité de résoudre, à chaque application concrète, les équations qui donnent c et α, en faisant une partie du calcul une fois pour toutes au moyen de tableaux numériques.

En posant

$$\frac{A}{\varepsilon} = h, \qquad \frac{c}{\varepsilon} = g, \qquad \frac{\alpha}{A} = l, \qquad \frac{\lambda_r}{A} = m,$$

les équations en c et α deviennent

$$rg^{r+1} - (r+1)g^r + h^{r+1} = 0, \tag{144}$$

$$l^{r+1} - (r+1)\,m^r\,(l-1) - 1 = 0 \tag{145}$$

Les données sont ε, A, λ_r d'où l'on déduit, par simples divisions, h et m. Pour chaque valeur de h (r étant choisi), un tableau numérique donnera par interpolation l'unique racine $> h$ de l'équation en g (144). La formule (130) devenant ici

$$p(\varepsilon) \leqq \left(\frac{mh}{g}\right)^r$$

fournira la borne cherchée.

Dans le cas où $m > 1$, on aura, pour ε suffisamment voisin de A, l'inégalité plus avantageuse

$$p(\varepsilon) \leqq \frac{l - \frac{1}{h}}{l-1}, \tag{146}$$

où l est l'unique racine > 1 de (145) et sera obtenue par interpolation au moyen d'un tableau numérique reliant les valeurs correspondantes de m et de l.

Pour s'assurer s'il y a lieu d'appliquer cette seconde inégalité, on vérifiera préalablement si

$$h > n,$$

où

$$n = \frac{r+1}{\frac{1}{l^r} + rl}$$

sera aussi fourni à partir de m au moyen d'un tableau numérique.

A titre d'exemple, dressons ces tableaux pour :

$$r = 2.$$

(147)

h ...	0	0,28	0,35	0,44	0,60	0,73	0,88	0,95	0,99	1
g	1,5	1,495	1,49	1,48	1,45	1,4	1,3	1,2	1,1	1

(147 *bis*)

m ..	1	1,53	2,08	2,65	3,22	6,00	8,96	13
n ..	1	0,70	0,49	0,37	0,29	0,14	0,09	0,06
l ...	1	2	3	4	5	10	15	22

Remarque. — Les formules (122) et (136) ci-dessus étant les plus avantageuses de leur espèce, il valait la peine d'en donner une démonstration complète. même longue. Mais une fois ces formules obtenues (Fréchet, **140**, p. 13, 14), il devenait souhaitable d'en trouver des preuves plus courtes.

Au moment de la correction des épreuves, M. von Mises m'informe qu'on peut en effet établir une démonstration moins longue en étendant convenablement, au cas de Camp généralisé, la méthode qu'il a employée dans son livre (**3**, p. 70) pour le cas de Gauss généralisé.

CHAPITRE V.

LES DIVERS MODES DE CONVERGENCE D'UNE SUITE DE VARIABLES ALÉATOIRES (¹)

SECTION I INTRODUCTION.

Premier point de départ. — On a souvent à considérer en Calcul des Probabilités des suites de variables aléatoires $X_1, X_2, \ldots X_n, \ldots$ et à se demander, pour chacune de ces suites, si elle est ou non convergente. Pour repondre à cette question, il faut d'abord s'entendre sur sa signification. *Il faut préciser ce qu'on entend par suite convergente de variables aléatoires.*

La maniere la plus naturelle (mais non la meilleure) de répondre à cette question est évidemment la suivante. On dira que la variable aléatoire X_n converge vers le nombre certain a quand n croît, si pour chaque épreuve déterminant l'ensemble des valeurs des X_n, la suite des valeurs déterminées prises par les X_n converge au sens arithmétique ordinaire vers a. Il est clair qu'une telle définition peut se généraliser au cas où la limite est elle-même une variable aléatoire X Elle suppose d'ailleurs que les X_n et X sont définis sur la même catégorie d'épreuves.

Un exemple simple est celui où X_n est la somme de la fraction $\frac{1}{n}$ et du nombre X de points obtenus en jetant un dé.

Mais le théorème de Bernoulli est venu dès les débuts du Calcul des Probabilités donner un exemple de convergence d'une nature différente. Soit X_n la fréquence $\frac{r}{n}$ d'un événement E de probabilité

(¹) Dans le présent Chapitre, les sections I, II, III, sont, à quelques modifications et additions près, la reproduction d'un mémoire paru en 1930 Fréchet, (135).

constante p au cours des n épreuves (appartenant à une même catégorie). Soient, d'autre part, ε un nombre positif arbitraire et ϖ_n la probabilité pour que $|X_n - p| > \varepsilon$.

Le théorème de Bernoulli ne tend pas à démontrer que, pour chaque suite infinie d'épreuves de c, la suite correspondante des valeurs bien déterminées des fréquences X_n tend vers p. Il affirme seulement que le nombre certain ϖ_n tend vers zéro avec $\frac{1}{n}$, c'est-à-dire qu'il est extrêmement peu probable, quand n est assez grand, que X_n diffère sensiblement de p, ou qu'il est très probable quand n est assez grand, que X_n diffère peu de p. Il y a là une notion tout à fait nouvelle et qu'on retrouve ensuite dans de multiples problèmes classiques du Calcul des Probabilités.

C'est ce dont s'était bien rendu compte Laplace. On peut le voir dans le passage suivant (*Œuvres*, vol. 10, p. 308) (signalé par M. Molina) « On voit ainsi comment les événements, en se multipliant, nous découvrent leur possibilité respective, mais on doit observer qu'il y a dans cette analyse deux approximations, dont l'une, est relative aux limites qui comprennent la valeur de $\mathfrak{X}$ et qui se resserrent de plus en plus et dont l'autre est relative à la probabilité que $\mathfrak{X}$ se trouve entre ces limites, probabilité qui approche sans cesse de l'unité ou de la certitude. C'est en cela que ces approximations diffèrent des approximations ordinaires dans lesquelles on est toujours assuré que le résultat est compris entre les limites qu'on lui assigne. »

Ainsi, l'énoncé même du théorème de Bernoulli conduit à une conception nouvelle de la méthode de convergence. D'une part, on considère une suite composée, non de nombres ou de fonctions, mais de valeurs aléatoires. D'autre part, au lieu d'imposer à ces valeurs aléatoires de tendre vers leur limite quelle que soit l'épreuve considérée, on se contente de s'assurer que la différence est probablement petite en précisant cette idée d'une façon convenable (nous verrons d'ailleurs plus loin, pages 164, 205, 208, ..., qu'on peut le faire de plusieurs façons non équivalentes et pourtant chacune assez naturelle).

C'est ce qui a amené M. Cantelli (3) à définir ce qu'il appelle la convergence « au sens du Calcul des Probabilités » d'une suite de variables aléatoires vers un nombre *certain*, et à en donner les propriétés.

Le cas d'une limite certaine, étudié par M. Cantelli, est en effet le

cas qui se présente dans le théorème de Bernoulli. Mais dans bien des questions qui se présentent naturellement, le cas d'une limite aléatoire doit être envisagé (nous en donnerons plus loin des exemples). C'est ce qui a été fait dans certains cas particuliers par M. Dell' Agnola (1. p. 57).

On pourrait ramener ce cas plus général au cas de M. Cantelli en convenant de dire que la variable aléatoire X_n converge « au sens du Calcul des Probabilités » vers la variable aléatoire X, quand n croît indéfiniment, si la variable aléatoire $X_n - X$ converge au sens du Calcul des Probabilités et de M. Cantelli. vers le nombre certain zéro. Mais d'abord, nous rappelons que, même pour le cas d'une limite certaine, il y a plusieurs façons non équivalentes (p. 161) d'interpréter la nouvelle espèce de limite. D'autre part, même si l'on se restreint à cette définition de la limite, il y a un certain nombre de questions qui se posent et où n'interviennent pas seulement $X_n - X$, mais X_n et X séparément. Sur le terrain plus uni de l'Analyse classique, on réduirait ou on compliquerait considérablement la théorie des limites de fonctions $f_n(x)$ si l'on n'étudiait jamais séparément le comportement de $f_n(x)$ et de sa limite $f(x)$, si l'on n'envisageait que leur différence $f_n(x) - f(x)$ pour ne considérer que le cas de limites constantes.

C'est à une extension du même ordre qu'est due l'intéressante conception de M. Slutsky (1, déf. 1, p. 40), celle d'une fonction certaine a_n, de n, stocastiquement asymptote à la suite de variables aléatoires X_n. Nous la généraliserons plus loin (p. 264) sous la forme de suites asymptotes de variables aléatoires X_n, Y_n, entendant par là, qu'il est très peu probable quand n est assez grand que X_n et Y_n different sensiblement.

Avant MM. Cantelli et Slutsky, M. Borel avait eu l'occasion, au sujet d'un problème de probabilité arithmétique, d'introduire une espèce de limite plus stricte que nous appellerons limite « presque certaine ».

Second point de départ. — On peut être conduit à ces mêmes notions *et à d'autres*, connexes, mais distinctes, par une toute autre voie, où nous sommes entré d'abord indépendamment, et dans l'ignorance, des travaux antérieurs de MM. Cantelli et Slutsky.

Au moment où un enseignement nouveau nous amenait à concentrer

nos pensées beaucoup plus qu'antérieurement, sur le Calcul des Probabilités, nous avons été tout naturellement conduit à *transposer dans le langage des Probabilités*, avec les modifications et les précautions convenables, un de nos précédents mémoires : *Sur divers modes de convergence d'une suite de fonctions d'une variable* [1]. Celui-ci, consacré à une question de *pure analyse*, avait été rédigé pour comparer entre elles diverses façons modernes d'envisager la notion de convergence. En remplaçant chaque fonction numérique $f(x)$ d'un nombre x par une fonction numérique X_E du résultat aléatoire E d'une épreuve, et en faisant jouer à la probabilité le rôle de la mesure linéaire, nous avons été naturellement conduit à étendre au Calcul des Probabilité les notions de convergence « en mesure » [2], de convergence « presque partout », de convergence uniforme « presque partout » qui avaient été comparées entre elles dans notre mémoire de Calcutta [1]. Nous avons aussi été amené à étendre la notion de « distance » de deux fonctions mesurables présentée dans ce même mémoire et à définir aussi la « distance » de deux variables aléatoires.

Nous avons tenu à marquer l'origine de ces extensions en introduisant l'expression de convergence « en probabilité » qui (tout en abrégeant utilement l'expression due a M. Cantelli de convergence « au sens du Calcul des Probabilités ») correspond à la convergence « en mesure ». De même, nous introduisons la notion de variables aléatoires « presque certainement » égales, en correspondance avec la notion due à M. Lebesgue, de fonctions égales « presque partout », etc.

Ce sont là des extensions *presque inéluctables* de la théorie moderne des fonctions de variables réelles

On pourrait aussi plus généralement étudier la convergence « au sens du Calcul des Probabilités » d'un élément aléatoire *de nature quelconque* X_E^n vers un élément aleatoire X_E de même nature, tous deux étant parfaitement définis par le résultat aléatoire E d'une épreuve. C'est ce qui a été amorcé dans le cas ou X_E^n, X_E sont deux

---

(1) *Bull Calcutta Math. Soc*, 1921 (vol. II de 1921), p 187-206.

(2) On sait que la convergence « en mesure » a été considérée à la fois par plusieurs auteurs : M F. Riesz, M Hardy (qui lui avait donné le nom de convergence asymptotique), etc

courbes (ou deux fonctions) et où E est un ensemble de valeurs numériques aléatoires en nombre fini fixe, par M. Glivenko (1).

Quelques inégalités utiles. Il nous sera commode par la suite d'établir quelques résultats préliminaires. Désignons par H.K l'événement consistant dans le concours des événements fortuits H et K, c'est-à-dire dans leur réalisation lors d'une même épreuve et par H + K la réalisation de l'un au moins des événements H. K. Rappelons qu'on a

$$\text{Pr. H} + \text{Pr. K} = \text{Pr.}[\text{H} + \text{K}] + \text{Pr.}[\text{H.K}] \tag{148}$$

On en déduit en particulier les inégalités qui nous seront utiles

$$\text{Pr.}[\text{H} + \text{K}] \leqq \text{Pr. H} + \text{Pr. K} \leqq 1 + \text{Pr.}[\text{H} + \text{K}], \tag{149}$$

$$\text{Pr.}[\text{H.K}] \leqq \text{Pr. H} + \text{Pr. K} \leqq 1 + \text{Pr.}[\text{H.K}] \tag{150}$$

En considérant d'abord pour simplifier le cas de deux variables aléatoires Y et Z et de deux nombres certains, A, B, appelons L, H, K les événements consistant en ce que

$$|Y + Z| \geqq A + B, \qquad |Y| \geqq A, \qquad |Z| \geqq B.$$

L'événement contraire à H + K consiste en ce que l'on ait simultanément

$$|Y| < A, \qquad |Z| < B, \qquad \text{d'où} \qquad |Y + Z| < A + B$$

Donc si L a lieu, H + K aussi et, par suite, en vertu de (149),

$$\text{Pr. L} \leqq \text{Pr.}[\text{H} + \text{K}] \leqq \text{Pr. H} + \text{Pr. K}.$$

Par voie de récurrence, on étend à n variables ([1]) le résultat obtenu.

LEMME. — *Si* $X_1, X_2, \ldots, X_n$ *sont* n *variables aléatoires définies sur la même catégorie d'épreuves et si* $A_1, \ldots, A_n$ *sont autant de nombres certains, la probabilité que*

$$|X_1 + \ldots + X_n| \geqq A_1 + A_2 + \ldots + A_n$$

est au plus égale à la somme de la probabilité que $|X_1| \geqq A_1, \ldots,$ *et de la probabilité que* $|X_n| \geqq A_n$.

Un raisonnement analogue montrerait que l'on a

$$\boxed{\begin{aligned}&\text{Pr.}[|X_1 + \ldots + X_n| > A_1 + \ldots + A_n]\\ &\leqq \text{Pr.}[|X_1| > A_1] + \ldots + \text{Pr.}[|X_n| > A_n]\end{aligned}} \tag{151}$$

([1]) M. Cramér observe que ce lemme résulte plus brièvement du fait que si $|X_1 + \ldots + X_n| > A_1 + \ldots + A_n$, l'un au moins des X_k est, en valeur absolue, supérieur à A_k.

et aussi

$$\boxed{\begin{aligned}&\Pr [X_1 + \ldots + X_n > A_1 + \ldots + A_n] \\ &\leqq \Pr [X_1 > A_1] + \ldots + \Pr [X_n > A_n].\end{aligned}} \qquad (152)$$

Dans ces inégalités les A sont chacun > 0, < 0 ou $= 0$.

Revenons à l'inégalité (150). Il est clair que si H est tres probable il sera presque aussi probable de voir se produire le concours de H et de K que de voir se produire K (avec ou sans H) Précisons : pour deux événements fortuits quelconques H et K, si

$$\Pr H \geqq 1 - \varepsilon,$$

on a

$$\Pr [H K] \geqq \Pr. K - \varepsilon. \qquad (153)$$

Cela résulte immédiatement de (150) et de l'hypothèse.

Application. — Supposons que Pr. H soit la probabilité que $|X - Y| < \eta$ et que Pr. K soit la probabilité que $Y < A$, X et Y étant deux valeurs aléatoires, η et A deux nombres certains. Si l'on a encore

$$\Pr. H \geqq 1 - \varepsilon,$$

on aura

$$\Pr (H K) \geqq \Pr K - \varepsilon$$

Or quand l'événement H.K a lieu, on a

$$X = (X - Y) + Y < A + \eta$$

Si donc L est l'événement consistant en ce que $X < A + \eta$, on aura

$$\Pr L \geqq \Pr. (H K) \geqq \Pr. K - \varepsilon$$

De même si L′ est l'événement $X \geqq A - \eta$, L′ est une conséquence des événements H et C(K) (contraire de K) et l'on aura

$$\Pr. L' \geqq \Pr. C(K) - \varepsilon;$$

d'où

$$\Pr. C(L') \leqq \Pr K + \varepsilon.$$

Ainsi, lorsque la probabilité de $|X - Y| \geqq \eta$ est $\leqq \varepsilon$, la probabilité que $Y < A$ est inférieure ou égale à la somme de ε et de la probabilité que $X < A + \eta$ est supérieure ou égale à l'excès sur ε de la probabilité que $X < A - \eta$.

Si $F_X(x)$ est la probabilité que $X < x$, et $F_Y(A)$ la probabilité que $Y < A$, on aura

$$F_X(A-\eta) - \varepsilon \leqq F_Y(A) \leqq \varepsilon + F_X(A+\eta).$$

On a d'ailleurs évidemment

$$F_X(A-\eta) \leqq F_X(A) \leqq F_X(A+\eta)$$

De ces deux systèmes d'inégalités, on déduit

$$|F_Y(A) - F_X(A)| \leqq \varepsilon + |F_X(A+\eta) - F_X(A-\eta)|$$

Nous arrivons ainsi à la proposition que nous avions en vue :

Lemme. — *Soient* X *et* Y *deux variables aléatoires définies sur la même catégorie d'épreuves.* A *et* η *deux nombres certains dont le second,* η, *est positif. On a alors, entre les fonctions des probabilités totales* F_X *et* F_Y, *de* X *et de* Y, *la relation*

$$|F_Y(A) - F_X(A)| \leqq |F_X(A+\eta) - F_X(A-\eta)| + \mathrm{Pr}\,[|X - Y| \geqq \eta] \quad (^1)$$

SECTION II CONVERGENCE « EN PROBABILITÉ »

Définitions et propriétés. — Soient X_n et X deux variables aléatoires définies sur la même catégorie d'épreuves. D'après la terminologie de M. Cantelli, la variable aléatoire $X_n - X$ converge « au sens du Calcul des Probabilités » vers le nombre certain zéro lorsque pour tout nombre positif η la probabilité $\mathrm{Pr.}[|X_n - X| < \eta]$ tend vers l'unité lorsque n croît indéfiniment

Pour les raisons exposées dans la seconde Introduction, page 160, nous formulerons une définition entièrement équivalente dans les termes suivants :

La variable aléatoire X_n *converge* (*ou tend*) « *en probabilité* » *vers la variable aléatoire* X (*lorsque n croît indéfiniment*), *si, pour tout nombre positif* η, *la probabilité que* X_n *diffère de* X *d'au moins* η *en valeur absolue, tend vers zéro avec* $\frac{1}{n}$. La condition de convergence « en probabilité » peut encore s'exprimer ainsi :

---

(1) Suivant M Cramér, ce lemme peut aussi se démontrer géométriquement en calculant de deux façons différentes les probabilités pour que le point X, Y appartienne aux domaines ($X < A$, $Y \geqq A$) et ($X \geqq A$, $Y < A$).

« ... si, pour tout couple de nombres positifs ε et η la probabilité que $|X_n - X| \geqq \eta$ reste inférieure à ε à partir d'un certain rang »

M. Cantelli a montré, et nous vérifierons plus loin, que si X_n converge « au sens du calcul des Probabilités » aussi bien vers le nombre *certain* X que vers le nombre *certain* Y, ces deux nombres sont nécessairement egaux. *Il n'en n'est plus de même dans le cas général*, où l'un au moins des deux nombres X et Y n'est pas supposé certain. Mais nous allons voir que les valeurs aléatoires X et Y sont intimement liées.

Par hypothèse, pour tout couple de nombres positifs ε, η, on a, à partir d'un certain rang N,

$$\Pr\left[|X - X_n| \geqq \frac{\eta}{2}\right] < \frac{\varepsilon}{2}, \qquad \Pr\left[|X_n - Y| \geqq \frac{\eta}{2}\right] \quad ,$$

En vertu du lemme de la page 162, on aura donc

$$\Pr\,[|X - Y| \geqq \eta] < \varepsilon \tag{154}$$

pour $n > N$

En posant $Z = |X - Y|$ et en désignant par $F_Z(x)$ la fonction de probabilité totale de Z, on voit qu'on a

$$1 - F_Z(\eta) < \varepsilon \tag{155}$$

Comme l'inégalité (155) est indépendante de n, l'inégalité (154) a lieu, sans condition, pour tout systeme de nombres positifs ε et η. On a donc, d'abord

$$1 - F_Z(\eta) \leqq 0$$

pour tout nombre positif η et, puisque $F_Z(\eta)$ est une probabilité, $F_Z(\eta) = 1$.

Par suite, $F_Z(+0) = 1$. Or, comme Z ne peut être négatif, $F_Z(0) = 0$. Dès lors, la probabilité que $Z = 0$, étant égale, en vertu de la remarque de la page 32, au saut à droite de F_Z au point $x = 0$, est égale à 1. Ainsi, s'il y a deux valeurs aléatoires X et Y qui puissent être indifféremment considérées comme limite d'une même suite X_1, X_2, ... convergente « en probabilité », il y a une probabilité égale à celle de la certitude que ces deux valeurs soient égales. Si ces valeurs X. Y sont certaines, elles sont alors nécessairement égales et nous retrouvons le résultat de M. Cantelli. Il en serait encore de

même si X et Y étaient des valeurs aléatoires ne prenant chacune qu'un nombre fini de valeurs numériques déterminées par un jeu de hasard. Dans les jeux (nous entendons les jeux usuels; cartes, dés, ...), la probabilité d'un des événements à considérer n'est égale à zéro ou à l'unité que s'il y a impossibilité ou certitude. Mais dans le cas le plus général, il est parfaitement légitime de considérer une valeur aléatoire $Z = (X - Y)$ non assujettie à être nulle et dont cependant la probabilité qu'elle soit nulle soit égale à l'unité [1].

Nous dirons que deux valeurs aléatoires X, Y définies sur la même catégorie d'épreuves sont « *presque certainement* » *égales* (ou « presque sûrement égales ») lorsqu'il y a une probabilité nulle qu'elles soient numériquement différentes

Étant donnée une suite de valeurs aléatoires X_n qui converge « en probabilité » vers une valeur aléatoire X, la condition nécessaire et suffisante pour qu'une valeur aléatoire Y soit aussi la limite « en probabilité » de X_n est que X et Y soient « presque certainement » égales.

Nous avons vu que la condition est nécessaire. Réciproquement, supposons que Y est « presque toujours » égal à X. On a, d'après la page 162,

$$\text{Pr.}[|X_n - Y| \geqq \eta] \leqq \text{Pr.}\left[|X_n - X| \geqq \frac{\eta}{2}\right] + \text{Pr.}\left[|X - Y| \geqq \frac{\eta}{2}\right],$$

et comme la dernière probabilité est nulle, pour tout $\eta > 0$, on aura

$$\text{Pr}\ [|X_n - Y| \geqq \eta] \leqq \text{Pr}\left[|X_n - X| \geqq \frac{\eta}{2}\right].$$

Or le second membre tend vers zéro avec $\frac{1}{n}$, donc aussi le premier

Si deux valeurs aléatoires Z, T sont « presque certainement » égales, on peut toujours les considérer comme limites « en probabilité » d'une même suite de variables aléatoires U_n. Il suffit de prendre par exemple $U_n = Z$, quel que soit n.

Deux valeurs aléatoires Z, U, « presque certainement » égales à une

---

[1] Par exemple, dans le cas où la probabilité qu'un point M d'un segment AB soit situé sur A′B′ est proportionnelle à A′B′, la probabilité que M soit par exemple au milieu de AB, est nulle. Dans ce cas, prenons $X = Y = AM$, sauf lorsque M est au milieu de AB, cas où nous prendrons $X = AM$, $Y = 2\,AM$. Et prenons $Z = X - Y$.

troisieme V, sont « presque certainement » égales entre elles. Cela résulte de ce qu'en vertu du lemme de la page 162, on a

$$0 \leqq \Pr.[|Z-U|>0] \leqq \Pr[|Z-V|>0]+\Pr.[|V-U|>0]=0$$

Les remarques suivantes sont souvent utiles :

I. Si deux valeurs aléatoires X_n, Y_n définies sur la même catégorie d'épreuves sont telles qu'on ait toujours $|X_n| \leqq |Y_n|$ et si Y_n tend « en probabilité » vers zéro, il en est de même de X_n.

Car

$$\Pr.[|Y_n| \geqq \eta] \geqq \Pr[|X_n| \geqq \eta]$$

pour toute valeur de $\eta > 0$.

II. Si des nombres aléatoires X_n, Y_n, ..., T_n en nombre fini fixé r convergent « en probabilité » vers X, Y, ..., T, la somme $X_n+Y_n+\ldots+T_n$ converge « en probabilité » vers $X+Y+\ldots+T$. Ceci résulte de ce que, en vertu du lemme de la page 162

$$\Pr[|X_n+\ldots+T_n-X-\ldots-T| \geqq \varepsilon]$$
$$\leqq \Pr\left[|X_n-X| \geqq \frac{\varepsilon}{r}\right]+\ldots+\Pr\left[|T_n-T| \geqq \frac{\varepsilon}{r}\right].$$

Premier critère de convergence en probabilité. — On doit à M. Slutsky (2) une condition de convergence en probabilité qui rappelle le critère de convergence de Cauchy. (En introduisant plus loin la notion de distance de deux variables aléatoires, il nous sera possible d'obtenir un critere, p. 196, dont la forme est plus proche encore de celle de Cauchy. Nous indiquerons aussi, p. 175, un autre critere fondé sur l'utilisation de la fonction des probabilités totales.)

Supposons que X_n converge en probabilité vers X, alors pour ε et η positifs arbitraires donnés, on a

$$\Pr\left\{|X_n-X|>\frac{\varepsilon}{2}\right\}<\frac{\eta}{2}$$

pour n assez grand . $n \geqq N$. On aura de même

$$\Pr.\left\{|X_m-X|>\frac{\varepsilon}{2}\right\}<\frac{\eta}{2} \qquad \text{pour } m \geqq N.$$

En vertu du lemme (p. 162), on aura donc

$$\text{(156)} \qquad \Pr \{ |X_n - X_m| > \varepsilon \} < \eta$$

pour n et m à la fois $\geqq N$.

Réciproquement, supposons que pour tout couple ε, η, l'inégalité (156) soit vérifiée pour n et m supérieurs à un nombre N convenablement choisi. Il s'agit d'en déduire l'existence d'une variable aléatoire X vers laquelle X_n converge en probabilité. Pour cela, considérons une série convergente à termes positifs $\sum_p \varepsilon_p$. Alors il existe, par hypothese, un nombre N_p, tel que

$$\text{Pr.} \{ |X_n - X_m| > \varepsilon_p \} < \varepsilon_p \qquad \text{pour } n \geqq N_p,\ m \geqq N_p$$

On peut évidemment supposer $N_{p+1} > N_p$.

Dès lors, pour tout entier p,

$$\text{Pr.} \{ |X_{N_p} - X_{N_{p+1}}| \leqq \varepsilon_p \} > 1 - \varepsilon_p.$$

Et en vertu de l'inégalité de Boole (23), p. 25, la probabilité de l'événement e_p résultant du concours des inégalités

$$\text{(157)} \quad |X_{N_p} - X_{N_{p+1}}| \leqq \varepsilon_p, \qquad \ldots \qquad |X_{N_{p+q}} - X_{N_{p+q+1}}| \leqq \varepsilon_{p+q+1}, \ldots$$

sera $> 1 - \varepsilon_p - \varepsilon_{p+1} \ldots$

Mais si les inégalités (157) ont lieu simultanément, la somme

$$(X_{N_1} - X_{N_2}) + \ldots + (X_{N_{r-1}} - X_{N_r}) = X_1 - X_{N_r}$$

converge vers une limite, c'est-à-dire la suite des X_{N_r} converge. L'événement H consistant dans la convergence de la suite des X_{N_r} a donc une probabilité au moins égale à celle de e_p. Ainsi l'on a

$$1 \geqq \text{Pr. H} > 1 - (\varepsilon_p + \varepsilon_{p+1} + \ldots)$$

et comme H est indépendant de p, on a

$$\text{Pr. H} = 1$$

Finalement, il y a déjà une suite au moins $X_{N_1}, X_{N_2}, \ldots$, extraite de la suite des X_n qui converge vers une variable aléatoire X bien déterminée quand H a lieu. Prenons arbitrairement X égal, par exemple, à zéro, quand H n'a pas lieu; alors X est une variable aléatoire bien définie.

Mais il faut revenir à la suite des X_n. Quand e_p a lieu, X_{N_p} converge vers X et, d'après (157),

$$|X_{N_p} - X| \leqq \varepsilon_p + \varepsilon_{p+1} + \ldots$$

Dès lors la probabilité de cette inégalité est $> 1 - \varepsilon_p - \varepsilon_{p+1} - \ldots$. D'autre part, on a

$$\Pr\{|X_n - X_{N_p}| \leqq \varepsilon_p\} > 1 - \varepsilon_p \qquad \text{pour } n > N_p$$

On a donc [form. (13), p. 16]

$$\Pr.\{|X_n - X| \leqq 2\varepsilon_p + \varepsilon_{p+1} + \varepsilon_{p+2} + \ldots\} > 1 - 2\varepsilon_p - \varepsilon_{p+1} - \varepsilon_{p+2} - \ldots$$
$$\text{pour } n > N_p$$

Si donc ε et η sont des nombres positifs quelconques, si l'on prend p assez grand pour que $2\varepsilon_p + \varepsilon_{p+1} + \varepsilon_{p+2} + \ldots$ soit inférieur à ε et η, on voit qu'on aura

$$\Pr\{|X - X_n| < \varepsilon\} > 1 - \eta \qquad \text{pour } n > N_p$$

Ainsi la suite des X_n converge en probabilité vers une variable aléatoire X.

D'où le résultat de M. Slutsky : *la condition nécessaire et suffisante pour qu'une suite de variables aléatoires* $X_1, \ldots, X_n, \ldots$ *soit convergente en probabilité est que, pour tout nombre* $\varepsilon > 0$ *et tout nombre* $\eta > 0$, *il existe un entier* N *tel que*

$$\Pr\{|X_n - X_m| > \varepsilon\} < \eta$$

quand n et m *sont tous deux* $> N$

Limite de la suite des fonctions des probabilités totales. — Soient X_1, $X_2, \ldots, X_n, \ldots$ une suite de variables aléatoires qui convergent « en probabilité » vers X ; soient $F_n(x)$ et $F(x)$ les fonctions des probabilités totales de X_n et de X, c'est-à-dire les probabilités respectives pour que $X_n < x$ et pour que $X < x$. Étant donnés les deux nombres positifs arbitraires, ε et η, il existe par hypothèse un entier N, tel que

$$\Pr.[|X_n - X| \geqq \eta] < \varepsilon \qquad \text{pour } n > N.$$

En vertu du lemme de la page 164, on a donc

$$|F_n(x) - F(x)| \leqq [F(x+\eta) - F(x-\eta)] + \varepsilon.$$

Si donc x est un point où F est continue, on pourra choisir ε et η

de sorte que le second membre soit aussi petit que l'on veut, et alors il en sera de même du premier pour n assez grand. $F_n(x)$ *converge donc vers* $F(x)$ *en tout point de continuité de* $F(x)$. Or $F(x)$ étant monotone, $F(x)$ n'a qu'une suite dénombrable de points de discontinuité. Ainsi $F_n(x)$ converge vers $F(x)$ pour toute valeur de x, sauf, peut-être, en un ensemble dénombrable de points.

Ceci suffit pour déterminer $F(x)$ *connaissant seulement les* $F_n(x)$. En effet, les $F_n(x)$ étant compris entre zéro et l'unité, la plus grande des limites de $F_n(x)$ quand, x restant fixe, n croît indéfiniment, est bien déterminée. Appelons-la $\Psi(x)$; cette fonction est monotone et elle est égale à $F(x)$ sauf peut-être en un ensemble dénombrable de points. On a donc partout $F(x-0) = \Psi(x-0)$ et comme on a vu (page 32) que $F(x)$ est partout continue à gauche, on voit que $F(x)$ est finalement partout déterminée par la formule

$$F(x) = \Psi(x-0)$$

Considérons d'abord le cas particulier où la fonction $F(x)$ (déterminée, par exemple, de cette façon) est partout continue. Alors elle est aussi (p. 32) uniformément continue Se donnant seulement $\varepsilon > 0$, on peut déterminer η, indépendamment de x, de sorte que

$$F(x+\eta) - F(x-\eta) < \varepsilon$$

et, par suite,

$$|F_n(x) - F(x)| < 2\varepsilon \qquad \text{pour } n > N$$

Or N est indépendant de x. Donc : *si la fonction* $F(x)$ *des probabilités totales de* X *est partout continue, elle est partout la limite* uniforme *de la fonction* $F_n(x)$ *des probabilités totales de* X_n.

Considérons maintenant le cas où $F(x)$ n'est pas partout continue. Il peut arriver que $F(x)$ soit continue dans certains intervalles, en nombre fini ou non. Si elle n'est pas continue aux deux extrémités de l'un de ces intervalles, elle n'y est peut-être pas uniformément continue, mais elle l'est dans tout intervalle intérieur. Le raisonnement fait plus haut prouvera que $F(x)$ est la limite uniforme de $F_n(x)$ sur tout intervalle où $F(x)$ est uniformément continue.

Observons qu'en un point ξ de discontinuité de $F(x)$, la seule suite des $F_n(\xi)$ donne un renseignement sur $F(\xi)$. En effet, si x' et x'' sont deux points de continuité de $F(x)$, encadrant ξ, la plus grande et la plus petite des limites des $F_n(\xi)$ seront évidemment comprises entre

$F(x')$ et $F(x'')$ et par suite aussi entre $F(\xi - 0)$ et $F(\xi + 0)$. Ainsi, *aux points de discontinuité de* $F(x)$, *la suite des* $F_n(\xi)$ *ne converge peut-être pas vers* $F(\xi)$, *mais toutes les limites de cette suite sont comprises entre* $F(\xi - 0) = F(\xi)$ *et* $F(\xi + 0)$.

Il peut être intéressant d'observer qu'*on peut toujours tirer de la suite des* X_n *une suite dont les fonctions de probabilités totales convergent pour toute valeur de* x. En effet, soit $c_1, c_2, \ldots$ l'ensemble, nécessairement dénombrable, des points de discontinuités de $F(x)$. Considérons $F_n(c_1), F_n(c_2), \ldots$ comme les coordonnées d'un point M_n de l'espace (E_ω) [*voir* E. A. (¹), p. 81]. Comme ces coordonnées sont entre 0 et 1, l'ensemble des M_n est borné et l'on peut en tirer (E. A., p. 117) (¹) une suite convergente en donnant à n une certaine suite de valeurs $n_1, n_2, \ldots$. Pour cette suite de valeurs de n, la suite des $F_n(x)$ convergera quel que soit x. La limite $\Psi(x)$ de cette suite sera égale à $F(x)$ aux points de continuité de $F(x)$; elle sera comprise entre $F(x)$ et $F(x+0)$ aux points de discontinuité de $F(x)$.

Les raisonnements précédents n'infirment pas l'hypothèse que la suite initiale des $F_n(x)$ soit elle-même convergente partout, ni celle qu'une suite convenablement extraite de celle des $F_n(x)$ converge partout précisément vers $F(x)$. Nous allons indiquer *des exemples montrant qu'aucune de ces hypothèses n'est nécessairement satisfaite.*

Prenons $X_{2s-1} = \frac{1}{s}$, $X_{2s} = -\frac{1}{s}$, $X = 0$. Ce sont, pour chaque valeur de s, des nombres certains, de sorte que non seulement X_n converge en probabilité vers X, mais même X_n converge au sens ordinaire vers X.

Or on a

$$F_{2s-1}(x) = 0 \quad \text{si } x \leqq \frac{1}{s},$$

$$F_{2s-1}(x) = 1 \quad \text{si } x > \frac{1}{s},$$

$$F_{2s}(x) = 0 \quad \text{si } x \leqq -\frac{1}{s},$$

$$F_{2s}(x) = 1 \quad \text{si } x > -\frac{1}{s}.$$

(¹) Nous représentons dans la suite, par la notation E. A., notre livre *Les espaces abstraits..*, chez Gauthier-Villars, Paris, 1928.

Donc la suite des $F_n(x)$ converge bien, pour $x \neq 0$, mais elle prend alternativement les valeurs 0 et 1 pour $x = 0$. La première hypothèse n'est donc pas satisfaite ici.

Il est vrai qu'on peut extraire de ces $F_n(x)$ une suite [à savoir la suite des $F_{2s}(x)$] qui converge partout vers $F(x)$. Mais cela n'a pas toujours lieu non plus. Il suffit pour le voir de prendre $X_n = -\frac{1}{n}$. Alors la suite initiale des $F_n(x)$ est bien elle-même convergente quel que soit x. Mais sa limite pour $x = 0$ est $1 \neq F(0) = 0$. La seconde hypothèse n'est donc pas vérifiée dans le second exemple.

Enfin, si l'on appelle X, Y deux valeurs aléatoires « presque certainement » égales et si l'on prend $X_n = Y$ quel que soit n, on conclut de ce qui précède que les fonctions des probabilités totales de X et de Y sont égales en tout point de continuité de l'une d'elles. Comme elles sont continues à gauche, elles sont donc égales partout : *deux valeurs aléatoires « presque certainement » égales ont même fonction de probabilité totale.*

D'ailleurs, *la réciproque n'est pas vraie.* Soient Z et T deux valeurs aléatoires ne prenant que les valeurs zéro et un, mais de sorte que si l'une prend la valeur zéro, l'autre prend la valeur 1. Non seulement ces nombres Z et T ne seront pas « presque certainement » égaux mais ils seront toujours inégaux et $|Z - T| = 1$ ne s'approchera jamais de zéro. Pourtant, s'il y a égale probabilité des deux valeurs zéro et un pour le premier, il en sera de même pour le second et leurs fonctions des probabilités totales seront identiques.

Remarque. — L'inexactitude de cette réciproque s'oppose à l'adoption d'un point de vue formulé par un des maîtres du Calcul des Probabilités : M. von Mises (4, p. 157), lequel, en parlant des variables aléatoires, déclare : « En vérité, il ne s'agit pas d'une classe spéciale de variables indépendantes, mais plutôt d'une classe spéciale de fonctions, à savoir de fonctions de répartition », c'est-à-dire de fonctions des probabilités totales. « Tous les théorèmes déduits dans cet ordre d'idées sont des théorèmes concernant les fonctions représentant les répartitions dans certains collectifs. » En réalité, comme on vient de le voir, une variable aléatoire X détermine une fonction des probabilités totales $F(x)$, mais non réciproquement. Les $F(x)$ sont des fonctions numériques d'une variable numérique x et

restent dans le cadre de l'Analyse classique. Les X sont des fonctionnelles, c'est-à-dire des fonctions numériques dont l'argument est le résultat d'une épreuve, résultat qui appartient bien à une classe spéciale de variables indépendantes. On ne saurait donc substituer la classe des fonctions des probabilités totales à la classe plus générale des variables aléatoires.

Cas particulier — Dans le cas particulier où l'une, Y par exemple, des deux variables aléatoires X, Y est un nombre certain ϖ, alors si X a même fonction des probabilités totales que Y, il y a une probabilité $= 0$ que $X < x$ si $x < \varpi$ et une probabilité égale à 1 que $X < x$ si $x > \varpi$. Il y a donc une probabilité nulle que $x' < X < x''$ si $x' < x'' < \varpi$ ou si $\varpi < x' < x''$. Il y a donc une probabilité nulle que $X \neq \varpi$ et, par suite, dans ce cas, la réciproque est vraie : X est presque certainement égal a ϖ.

En particulier la réciproque est vraie pour deux nombres certains : si deux nombres certains ont même fonction de probabilité totale, ils sont égaux.

Signification de l'identité de deux fonctions des probabilités totales. — Il est d'ailleurs facile de voir ce qu'exprime l'identité des fonctions des probabilites totales de deux valeurs aléatoires quelconques X, Y. Appelons E_x l'événement consistant en ce que $X = x$ et F_x l'événement consistant en ce que $Y = x$ On voit que la certitude peut être répartie, soit entre les événements E_x correspondant à toutes les valeurs numériques de x (peut-être n'y aura-t-il aucun événement correspondant à certaines valeurs de x?), soit entre les événements F_x. Dire que les fonctions de probabilité totale des deux nombres aléatoires X et Y sont égales, c'est dire que ces deux répartitions (qui peuvent être très différentes) etant faites, un ensemble d'événements E_x et un ensemble d'événements F_x correspondant (c'est-à-dire avec les mêmes indices) ont même probabilité. Ou, tout au moins, si l'on veut éviter toute difficulté sur l'existence des probabilités, il doit en être ainsi quand chacun des ensembles a une probabilité définie.

Autrement dit, on obtient toute valeur aléatoire Y qui a même fonction de probabilité totale que X, en effectuant une transformation de l'ensemble des événements possibles E_x où X a une valeur constante x

en lui-même, de sorte que les probabilités soient conservées dans la transformation et en attribuant à Y pour un événement déterminé transformé de E, la valeur x qu'avait X avant la transformation.

Réciproque. — L'exemple des deux variables Z et T de la page 172 nous montre que : *si les fonctions des probabilités totales* $F_n(x)$ *d'une suite de variables aléatoires* X_n *convergent vers la fonction des probabilités totales* $F(x)$ *d'une variable aléatoire* X, *aux points de continuité de* $F(x)$, *on ne doit point, en général, en conclure que* X_n *converge en probabilité vers* X. Il suffit, en effet, d'observer que si l'on prend $X_n \equiv Z$ et $X \equiv T$, on aura

$$F_n(x) \equiv F(x),$$

quels que soient n et x, on a donc simultanément

$$F(x) = \lim_{n \to \infty} F_n(x)$$

et pourtant

$$|X_n - X| = 1$$

dans toute épreuve et quel que soit n.

Cependant, la conclusion contestée redevient exacte (Cantelli, 2, p. 278) quand X est un nombre certain ϖ.

En effet, dans ce cas, $F(x)$ étant égal à 0 pour $x < \varpi$ et à 1 pour $x > \varpi$, on a pour tout ε positif

$$(158) \quad 1 - \Pr\left\{ \varpi - \varepsilon < X_n \leqq \varpi + \varepsilon \right\} = [F(\varpi + \varepsilon) - F(\varpi - \varepsilon)] - [F_n(\varpi + \varepsilon) - F_n(\varpi - \varepsilon)]$$

En vertu de l'hypothèse, le second membre tend vers zéro avec $\frac{1}{n}$; donc aussi le premier et l'on a

$$(159) \quad \lim_{n \to \infty} \Pr\left\{ |X_n - \varpi| \leqq \varepsilon \right\} = 1 \qquad \text{pour tout } \varepsilon > 0 \quad [1].$$

C'est-à-dire que X_n converge en probabilité vers $\varpi = X$.

Second critère de la convergence en probabilité. — Il en résulte alors en remplaçant X_n par $Y_n - Y$ et ϖ par zéro que : *la condition*

---

[1] Cela n'est démontré que pour les valeurs de ε telles que $\varpi - \varepsilon$ soient des points de continuité de $F(x)$. Mais il en est alors ainsi nécessairement pour tout $\varepsilon > 0$, puisque le premier membre de (159) est une fonction monotone de ε.

nécessaire et suffisante pour qu'une suite de variables aléatoires Y_n *converge en probabilité vers la variable aléatoire* Y *est que la fonction des probabilités totales de* $Y_n - Y$, *soit* $\Phi_n(x)$, *converge vers zéro pour* $x < 0$ *et vers l'unité pour* $x > 0$ (Cantelli, 2, p. 276).

Convergence de la médiane. — Considérons d'abord le cas le plus simple, celui où $F_n(x)$ et $F(x)$ passent en croissant par la valeur $\frac{1}{2}$ et où, par suite, X_n a une seule médiane m_n et X une seule médiane m. Il est alors à peu près évident que *si* X_n *converge en probabilité vers* X, m_n *converge vers* m. Car, dans le cas contraire, il y aurait un nombre positif ε tel que m_n fût en dehors de l'intervalle $m - \varepsilon$, $m + \varepsilon$ pour une infinité de valeurs de n. Soit ε_0 un nombre positif $< \varepsilon$, tel que $F(x)$ soit continu pour $x = m - \varepsilon_0$, et pour $x = m + \varepsilon_0$. On aurait, pour une infinité de valeurs de n,

$$\frac{1}{2} = F_n(m_n) \leqq F_n(m - \varepsilon_0) \qquad \text{ou} \qquad F_n(m + \varepsilon_0) \leqq F_n(m_n) = \frac{1}{2}.$$

A la limite, on aurait donc

$$\frac{1}{2} \leqq F(m - \varepsilon_0) < F(m) = \frac{1}{2} \qquad \text{ou} \qquad \frac{1}{2} = F(m) < F(m + \varepsilon_0) \leqq \frac{1}{2},$$

ce qui est impossible.

On peut étendre ce résultat, dû à M. Slutsky (1), et précédé par un résultat plus particulier mais analogue, dû à M. Cantelli (4, p. 344-345), au cas général où il peut y avoir pour X_n et X une infinité de valeurs médianes. Une valeur médiane m_n de X_n est tout nombre m_n tel que

$$F_n(m_n - 0) \leqq \frac{1}{2} \leqq F(m_n + 0),$$

c'est-à-dire tel que

$$F_n(x) \leqq \frac{1}{2} \qquad \text{pour } x < m_n$$

et

$$F_n(x) \geqq \frac{1}{2} \qquad \text{pour } x > m_n.$$

Quand n croît, m_n peut avoir plusieurs limites (ou même une infinité). Soit m l'une des limites de m_n, de sorte que $m = \lim_{j \to \infty} m_{n_j}$,

nous voulons prouver que m est une médiane de X. Soit $x < m$. Pour j assez grand on aura

$$m_{n_j} > x,$$

d'où

$$F_{n_j}(x) \leqq F_{n_j}(m_{n_j}) \leqq \frac{1}{2}.$$

Si l'on a pris pour x un point de continuité de $F(x)$; on aura

$$F(x) = \lim_{j \to \infty} F_{n_j}(x) \leqq \frac{1}{2}.$$

Ainsi pour tout point x de continuité de F à gauche de m, on a

$$F(x) \leqq \frac{1}{2}; \qquad \text{d'où} \qquad F(x - 0) \leqq \frac{1}{2}.$$

On verrait de même que

$$F(x + 0) \geqq \frac{1}{2}.$$

En résumé, *quand* X_n *converge en probabilité vers* X, *toutes les limites, quand n croît de toutes les médianes de* X_n, *sont des médianes de* X.

Comportements divers de la moyenne. — Par contre, M. Slutsky (1) a observé que, dans la convergence en probabilité de X_n vers X, le comportement de la moyenne V_n de X_n n'est pas aussi simple que celui de la médiane.

Par exemple. considérons avec M. Slutsky le cas où X_n ne prend que les valeurs $-n$, -1, n avec les probabilités $\frac{1}{n}$, $1 - \frac{4}{n}$, $\frac{3}{n}$. Alors il est visible que X_n converge en probabilité vers la constante $X = -1$. Or

$$V_n = \frac{1}{n}(-n) + \left(1 - \frac{4}{n}\right)(-1) + \frac{3}{n}(n) = 1 + \frac{4}{n};$$

donc

$$V_n \to 1 \neq V = -1$$

Ainsi : *quand* X_n *converge en probabilité vers* X, *la valeur moyenne de* X_n *peut ne pas converger vers celle de* X.

Il y a pourtant un cas simple et important où cette singularité ne peut se présenter : c'est, comme il résultera de l'étude plus générale de la page 188, le cas où les X_n et X restent, en toute épreuve,

compris entre deux nombres fixes indépendants de l'épreuve et de n. C'est aussi, plus généralement, le cas où, pour au moins une valeur convenable de s, l'écart moyen d'ordre s de X_n et X reste borné à partir d'une valeur assez grande de n. Il résultera de la page 188 que, dans ce cas, les écarts moyens de tous ordres $< s$ tendent vers zéro.

Pour que cette valeur de s soit convenable, il suffit que $s > 1$. Alors l'écart d'ordre 1 tendra vers zéro et il suffira d'utiliser l'inégalité évidente

$$|\mathfrak{M}X_n - \mathfrak{M}X| \leqq \mathfrak{M}|X_n - X|$$

pour voir que $\mathfrak{M}X_n$ tend vers $\mathfrak{M}X$; on vérifie en même temps une observation supplémentaire de M. Slutsky, à savoir que, dans le cas où $s > 2$, l'écart quadratique moyen de X_n et X tend aussi vers zéro.

Convergence « en probabilité » des fonctions continues. — Faisons d'abord quelques remarques préparatoires. Soient $q_n(t)$, $q(t)$ les probabilités pour que $|X_n| < t$; $|X| < t$. Si X_n tend « en probabilité » vers X, $|X_n|$ tend aussi « en probabilité » vers $|X|$, car

$$|X_n - X| \geqq \big||X_n| - |X|\big|$$

Donc $q_n(t)$ tend vers $q(t)$ aux points de continuité de $q(t)$

Nous supposons que X_n et X sont toujours finis (non nécessairement bornés). Par suite, $q_n(t)$ et $q(t)$ tendent vers l'unité quand t croît indéfiniment. Pour tout entier n et tout $\varepsilon > 0$ on peut donc fixer B_n tel que

$$q_n(B_n) > 1 - \varepsilon$$

Lemme. — *On peut choisir* B_n *indépendant de n.*

Pour cela prenons un nombre B_0 assez grand pour que

$$q(B_0) > 1 - \frac{\varepsilon}{2}.$$

On peut supposer que B_0 soit un point de continuité de $q(t)$. Alors en prenant n assez grand ($n > N$), on aura

$$q_n(B_0) - q(B_0) > -\frac{\varepsilon}{2};$$

d'où

$$q_n(B_0) > 1 - \varepsilon$$

Prenons pour B le plus grand des nombres $B_0, B_1, \ldots, B_N$. On voit qu'on aura, pour tout entier n,

$$q_n(B) > 1 - \varepsilon$$

Ceci étant, soit $\varphi(x)$ une fonction continue de x, elle est uniformément continue pour $|x| \leqq B$. Si ω est un nombre positif donné en même temps que ε, il existe un nombre η tel que $|\varphi(x) - \varphi(x')| < \omega$ quand x, x' varient dans $(-B, +B)$ sous la condition $|x - x'| < \eta$

Si l'on a simultanément, pour une même épreuve,

$$|X| < B, \qquad |X_n| < B, \qquad |X - X_n| < \eta,$$

on aura

$$|\varphi(X) - \varphi(X_n)| < \omega$$

Ces trois événements ont des probabilités chacune au moins égale à $1 - \varepsilon$, pour des valeurs de n assez grandes ($n > M$) d'après ce qui précède et puisque la probabilité de l'inégalité $|X - X_n| < \eta$ tend vers l'unité quand n croît indéfiniment. Le concours de ces trois événements a donc, pour $n > N$, une probabilité supérieure à $1 - 3\varepsilon$. Alors la probabilité que

$$|\varphi(X) - \varphi(X_n)| < \omega$$

est *a fortiori* supérieure à $1 - 3\varepsilon$. Finalement, il est démontré que *si $\varphi(x)$ est une fonction continue de x et si X_n est une variable aléatoire qui tend « en probabilité » vers la variable aléatoire X, $\varphi(X_n)$ est une variable aléatoire qui tend « en probabilité » vers la variable aléatoire $\varphi(X)$.*

Ce théorème avait été déjà établi pour le cas où X est certain par M. Slutsky (1, p. 76, th. 18), puis étendu par lui, sous une restriction que nous avons pu négliger, au cas où X est aléatoire (Slutsky, 1, th. I, II).

Cas de plusieurs variables. — *Définition.* — Un point aléatoire est un point dont la position est déterminée par le résultat d'une épreuve, ce résultat étant lui-même déterminé par le hasard. Nous dirons qu'une suite de points $M_1, M_2, \ldots$ tend « en probabilité » vers un point M, si l'on a

$$\lim_{n \to \infty} P_n(\varepsilon) = 0,$$

en désignant par $P_n(\varepsilon)$ la probabilité que la distance $M_nM \geqq \varepsilon$.

THÉORÈME. — *La condition nécessaire et suffisante pour que le point aléatoire* M_n *converge « en probabilité » vers le point aléatoire* M *est que chacune des coordonnées de* M_n *converge « en probabilité » vers la coordonnée correspondante de* M.

Supposons, par exemple, que M_n et M restent dans un plan.
Soient X_n, Y_n les coordonnées de M_n, X, Y celle de M. On a

$$|X_n - X| \leqq M_nM \quad \text{et} \quad |Y_n - Y| \leqq M_nM$$

Donc la probabilité que $M_nM \geqq \varepsilon$ est au moins égale à la probabilité que $|X_n - X| \geqq \varepsilon$. Or si M_n tend « en probabilité » vers M, la premiere probabilité tend vers zéro, donc l'autre aussi, donc X_n tend vers X, « en probabilité », et de même Y_n tend vers Y « en probabilité ». D'autre part, soient $p_n(\varepsilon)$, $p'_n(\eta)$ les probabilités respectives que $|X_n - X| \geqq \varepsilon$, $|Y_n - Y| \geqq \eta$. On a

$$M_nM \leqq |X_n - X| + |Y_n - Y|$$

Donc la probabilite que $M_nM \geqq \varepsilon + \eta$ est au plus égale à la probabilité que

$$|X_n - X| + |Y_n - Y| \geqq \varepsilon + \eta$$

et celle-ci est au plus égale, d'apres la page 162, à $p_n(\varepsilon) + p'_n(\eta)$. Donc, si X_n et Y_n convergent « en probabilité » vers X et Y, M_n converge « en probabilité » vers M.

Il est clair que le raisonnement s'étendrait immédiatement au cas où M, M_n seraient deux points d'un même espace cartésien à un nombre fini quelconque de dimensions [1]. La proposition avait déjà été établie dans le cas où M est un point certain par M. Slutsky (**1**, p. 61, th. **12** et **13**).

LEMME. — Soient O un point fixe, M et M_n deux points aléatoires, M_n convergeant « en probabilité » vers M; posons

$$q(R) = \text{Pr.}[OM < R], \qquad q_n(R) = \text{Pr.}[OM_n < R]$$

Alors pour tout $\varepsilon > 0$, il y a un nombre ρ indépendant de n tel qu'on

---

[1] Il n'en serait pas de même dans chacun des espaces à une infinité de dimensions dont la considération s'est imposée.

ait à la fois

$$q(\rho) > 1 - \varepsilon, \qquad q_n(\rho) > 1 - \varepsilon$$

quel que soit n.

Il suffit, pour le voir, d'observer que OM et OM_n sont deux valeurs aléatoires et d'appliquer le lemme de la page 177, en posant $X = OM$, $X_n = OM_n$, $t = R$, $B = \rho$.

Théorème. — *Si $\varpi(x, y)$ est une fonction continue de l'ensemble des variables x, y et si X_n, Y_n sont deux variables aléatoires qui tendent « en probabilité » la première vers X, la seconde vers Y, alors $\varpi(X_n, Y_n)$ est une variable aléatoire qui converge « en probabilité » vers la variable aléatoire $\varpi(X, Y)$.*

Il suffit de refaire le raisonnement détaillé pour une variable en prenant ici comme variable le point aléatoire M_n de coordonnées X_n, Y_n, et en considérant $\varpi(X, Y)$ comme une fonction $u(M)$ du point aléatoire M de coordonnées X, Y.

Remarques. — 1° Il est clair que la démonstration et le résultat précédents s'étendent à un nombre fini quelconque de variables.

2° Si les nombres aléatoires X et X_n restent compris entre deux nombres fixes et s'il en est de même de Y et Y_n, le théorème restera exact quand la continuité de $\varpi(x, y)$ n'est admise que dans le rectangle correspondant.

3° Si X, Y sont des nombres certains, il suffit, pour assurer le dernier résultat énoncé, de supposer que $\varpi(x, y)$ est continue au seul point (X, Y). Ce cas particulier a été déjà énoncé et démontré par M. Cantelli pour un nombre fini quelconque de variables.

Valeur moyenne d'une fonction. — On va étudier la suite des valeurs moyennes d'une certaine fonction de x : $\varphi(x)$ lorsqu'on y remplace x par une variable aléatoire X_n qui convergera « en probabilité » vers la variable aléatoire X.

Il paraît naturel de prévoir qu'on n'arrivera à des résultats simples que si cette fonction elle-même est simple. Il est remarquable qu'en réalité, on puisse, pour y arriver, n'assujettir $\varphi(x)$ qu'à des conditions très générales : il nous suffira, dans ce qui suit, *de supposer que $\varphi(x)$ est continue.*

Désignons par $\overline{Y}$ la valeur moyenne d'une variable aléatoire Y. On a vu, page 55, que l'on a

$$\overline{\varphi(X_n)} = \int_{-\infty}^{+\infty} \varphi(x)\, d\,F_n(x)$$

si cette intégrale est absolument convergente.

Montrons d'abord que si a, b sont des points de continuité de $\varphi(x)$, on a

$$\int_a^b \varphi(x)\, d\,F(x) = \lim_{n \to \infty} \int_a^b \varphi(x)\, d\,F_n(x). \tag{160}$$

Puisque la variation totale de $F_n(x)$ est égale à l'unité, on a

$$\int_a^b \varphi(x)\, d\,F_n(x) = \Sigma\, \varphi(x_i)[F_n(x_{i+1}) - F_n(x_i)] + \theta_n \Omega,$$

en désignant par Ω l'oscillation de $\varphi(x)$ dans un intervalle de longueur δ égale au plus grand des intervalles (x_i, x_{i+1}) et en prenant $|\theta_n| \leqq 1$.

Si l'on prend les x_i parmi les points de continuité communs aux F_n et à F, les $F_n(x_i)$ tendent vers les $F(x_i)$ quand, les x_i restant fixes, n croît indéfiniment. Or, on a une égalité analogue pour $\int_a^b \varphi(x)\, d\,F(x)$ et l'on peut prendre Ω aussi petit que l'on veut.

L'égalité (160) est donc bien établie.

Considérons maintenant, d'abord, le cas simple où X et les X_n sont uniformément bornés, c'est-à-dire où il existe deux nombres fixes A, B entre lesquels X et X_n doivent rester compris quelle que soit l'épreuve qui les détermine et quel que soit le rang n. Alors $F(x)$ est constant en dehors de (A, B).

Par définition,

$$\int_{-\infty}^{+\infty} \varphi(x)\, d\,F(x) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b \varphi(x)\, d\,F(x)$$

Puisque l'intégrale du second membre est constante quand $a < A$ et $b > B$, on aura, sous ces hypothèses,

$$\overline{\varphi(X)} = \int_a^b \varphi(x)\, d\,F(x) = \lim_{n \to \infty} \int_a^b \varphi(x)\, d\,F_n(x) = \lim_{n \to \infty} \overline{\varphi(X_n)}.$$

Ainsi, $\varphi(x)$ étant une fonction continue arbitraire, il suffit que X et X_n soient uniformément bornés pour que, lorsque X_n tend « en probabilité » vers X, la valeur moyenne de $\varphi(X_n)$ tende vers la valeur moyenne de $\varphi(X)$.

Considérons le cas général où X et les X_n ne sont pas uniformément bornés.

En continuant à supposer que $\varphi(x)$ est une fonction continue, il est clair qu'on ne peut avoir

$$\overline{\varphi(X)} = \lim_{n \to \infty} \overline{\varphi(X_n)}$$

pour toute suite X_n convergeant « en probabilité » vers X. En effet, pour que cette relation soit exacte, il faut au moins qu'elle ait un sens. Si l'on suppose que $\overline{\varphi(X)}$ soit finie, il faut donc au moins supposer qu'à partir d'un certain rang n, $\overline{\varphi(X_n)}$ ait une valeur bien déterminée. C'est-à-dire que

$$\int_a^b \varphi(x)\, d\,F_n(x)$$

converge, pour chaque valeur fixe de n, vers une limite bien déterminée quand a et b tendent respectivement vers $-\infty$ et $+\infty$. C'est ce qui a lieu nécessairement dans le cas que nous venons d'examiner, où X_n restant borné, $F_n(x)$ est constant pour $|x|$ assez grand. Mais dans le cas général, la convergence de chaque intégrale définie est une nouvelle hypothèse.

Nous allons d'abord considérer le cas où *la convergence* de

$$\int_a^b \varphi(x)\, d\,F_n(x),$$

quand n restant fixe, a et b tendant respectivement vers $-\infty$ et $+\infty$, *est uniforme*, au moins à partir d'un certain rang. C'est-à-dire que (à partir du rang N où cette intégrale converge), pour tout $\varepsilon > 0$, on peut déterminer A et B tels que

$$\left| \int_a^b \varphi(x)\, d\,F_n(x) - \int_{-\infty}^{+\infty} \varphi(x)\, d\,F_n(x) \right| < \varepsilon \qquad \text{pour } a < A \text{ et } b > B,$$

A et B restant indépendants de n.

Puisqu'on suppose $\overline{\varphi(X)}$ fini, on peut supposer A et B assez grands

pour qu'on ait aussi

$$\left|\int_a^b \varphi(x)\,d\mathrm{F}(x) - \int_{-\infty}^{+\infty} \varphi(x)\,d\mathrm{F}(x)\right| < \varepsilon \qquad \text{pour } a < \mathrm{A} \text{ et } b > \mathrm{B}.$$

Or, d'après ce qui précède, lorsque l'on a ainsi fixé a et b, il y a aussi un rang N' tel que

$$\left|\int_a^b \varphi(x)\,d\mathrm{F}(x) - \int_a^b \varphi(x)\,d\mathrm{F}_n(x)\right| < \varepsilon.$$

On aura donc, pour $n > \mathrm{N} + \mathrm{N}'$,

$$|\overline{\varphi(\mathrm{X})} - \overline{\varphi(\mathrm{X}_n)}| = \left|\int_{-\infty}^{+\infty} \varphi(x)\,d\mathrm{F}(x) - \int_{-\infty}^{+\infty} \varphi(x)\,d\mathrm{F}_n(x)\right| < 3\varepsilon,$$

c'est-à-dire

$$\overline{\varphi(\mathrm{X})} = \lim_{n\to\infty} \overline{\varphi(\mathrm{X}_n)}$$

Inversement, supposons que l'on ait cette égalité. $\varphi(x)$ étant encore une fonction continue, les moyennes $\overline{\varphi(\mathrm{X}_n)}$ étant finies à partir d'un certain rang N, ainsi que $\overline{\varphi(\mathrm{X})}$ A partir de ce même rang N, on peut écrire

$$\left|\int_{-\infty}^{+\infty} \varphi(x)\,d\mathrm{F}_n(x) - \int_{a_0}^{b_0} \varphi(x)\,d\mathrm{F}_n(x)\right|$$
$$\leqq |\overline{\varphi(\mathrm{X}_n)} - \overline{\varphi(\mathrm{X})}| + \left|\int_{-\infty}^{+\infty} \varphi(x)\,d\mathrm{F}(x) - \int_{a_0}^{b_0} \varphi(x)\,d\mathrm{F}(x)\right|$$
$$+ \left|\int_{a_0}^{b_0} \varphi(x)\,d\mathrm{F}(x) - \int_{a_0}^{b_0} \varphi(x)\,d\mathrm{F}_n(x)\right|$$

Soit $\varepsilon > 0$. On peut prendre N', a_0, b_0 de sorte que les deux premiers termes du second membre soient $< \frac{\varepsilon}{3}$ pour $n > \mathrm{N}' > \mathrm{N}$. Mais a_0 et b_0 étant ainsi choisis, on pourra prendre $\mathrm{N}'' > \mathrm{N}'$, de sorte que le dernier terme du second membre soit $< \frac{\varepsilon}{3}$ pour $n > \mathrm{N}''$. De sorte qu'en prenant $\mathrm{N}''' = \mathrm{N}''' + \mathrm{N}'$, le premier membre sera, pour un choix particulier de valeurs de a_0 et b_0, inférieur à ε pour $n > \mathrm{N}'''$. D'autre part, il est clair qu'à partir du rang N [depuis lequel $\overline{\varphi(\mathrm{X}_n)}$

est fini]. on peut déterminer des nombres a_n, b_n tels que

$$\left| \int_{-\infty}^{+\infty} \varphi(x)\, d\,F_n(x) - \int_{a_n}^{b_n} \varphi(x)\, d\,F_n(x) \right| < \varepsilon$$

Soient maintenant. A un nombre inférieur à $a_0, a_{N+1}, \ldots, a_{N''}$ et B un nombre supérieur à b_0, b_{N+1}, ..., $b_{N''}$. Si nous faisons maintenant *sur* $\varphi(x)$ *une nouvelle hypothèse*, à savoir que $\varphi(x)$ est constamment $\geqq 0$, nous voyons qu'on aura

$$\left| \int_{-\infty}^{+\infty} \varphi(x)\, d\,F_n(x) - \int_{a}^{b} \varphi(x)\, d\,F_n(x) \right| < \varepsilon$$

à partir d'un rang N *indépendant* de ε, pour $a < A$ et $b > B$. Autrement dit, il est prouvé que si $\varphi(x) \geqq 0$, la condition de l'uniformité de la convergence de

$$\int_{a}^{b} \varphi(x)\, d\,F_n(x) \quad \text{vers} \quad \int_{-\infty}^{+\infty} \varphi(x)\, d\,F_n(x)$$

est non seulement suffisante mais *nécessaire* pour assurer l'égalité

$$\overline{\varphi(X)} = \lim_{n \to \infty} \overline{\varphi(X_n)}$$

Dans le cas où $\varphi(x)$ ne serait pas constamment positif ou nul, la condition serait encore nécessaire si l'on supposait qu'on ait non seulement la dernière égalité, mais encore

$$\overline{|\varphi(X)|} = \lim_{n \to \infty} \overline{|\varphi(X_n)|}$$

En effet, si $\varphi(x)$ est continue, $|\varphi(x)|$ l'est aussi; il y aurait donc, d'après ce qui précède, uniformité de la convergence des intégrales

$$\int_{a}^{b} |\varphi(x)|\, d\,F_n(x)$$

et par suite, *a fortiori*, uniformité de la convergence des intégrales

$$\int_{b}^{a} \varphi(x)\, d\,F_n(x).$$

Remarque. — Dans le cas où $\varphi(x)$ reste constamment positif ou nul, si la convergence des intégrales n'a pas lieu uniformément, on

peut cependant donner une indication sur la suite des $\overline{\varphi(X_n)}$, connaissant seulement $\overline{\varphi(X)}$. En effet, on a

$$\overline{\varphi(X_n)} = \int_{-\infty}^{+\infty} \varphi(x)\, d\,F_n(x) \geqq \int_a^b \varphi(x)\, d\,F_n(x).$$

Or, quand a et b sont deux nombres arbitraires fixes, le dernier terme tend vers

$$\int_a^b \varphi(x)\, d\,F(x)$$

Si donc ϖ est la plus petite des limites de $\overline{\varphi(X_n)}$ quand n croît indéfiniment, on aura quels que soient a et b

$$\varpi \geqq \int_a^b \varphi(x)\, d\,F(x)$$

et, par suite,

$$\varpi \geqq \int_{-\infty}^{+\infty} \varphi(x)\, d\,F(x)$$

Ainsi, lorsque $\varphi(x)$ reste $\geqq 0$, $\overline{\varphi(X)}$ est inférieure ou égale à la plus petite des limites, pour n infini, de $\overline{\varphi(X_n)}$. En particulier, si l'on connaît la suite des $\overline{\varphi(X_n)}$ sans connaître $\overline{\varphi(X)}$, cette proposition permet de conclure que $\overline{\varphi(X)}$ ne peut être infinie que si la suite des $\overline{\varphi(X_n)}$ tend vers l'infini (par valeurs finies ou non).

Cas des écarts moyens. — Si $X_1, X_2, \ldots$ est une suite de variables aléatoires qui tend « en probabilité » vers la variable aléatoire X, et si ω est un nombre certain, la condition nécessaire et suffisante pour que l'écart moyen d'ordre r de X_n avec ω tende vers l'écart moyen d'ordre r de X avec ω est que la convergence de

$$\int_a^b |x - \omega|^r\, d\,F_n(x)$$

vers la valeur moyenne de $|X - \omega|^r$ (quand r restant fixe, a et b tendent respectivement vers $-\infty$ et $+\infty$) soit uniforme quand n varie.

Cette condition d'uniformité de la convergence des intégrales,

sera, en particulier, remplie quand X_n et X sont uniformément bornes.

Plus généralement, si avec les notations précédentes, X_n converge en probabilité vers X, la condition nécessaire et suffisante pour que le « g-écart moyen » (défini page 120) de X_n avec ϖ converge vers le g-écart moyen de X avec ϖ, est que la convergence de l'intégrale $\int_a^b g(|X_n - \varpi|)\, d\,F_n(x)$ (quand a et b tendent respectivement vers $-\infty$ et $+\infty$) soit uniforme quand n varie pour n assez grand.

Il est bon de montrer par un exemple que si une suite de nombres aléatoires X_n converge « en probabilité » vers un nombre aléatoire X, l'écart moyen d'ordre r de X_n avec ϖ ne converge pas nécessairement vers celui de X avec ϖ.

Considérons une catégorie d'épreuves dans chacune desquelles se produise nécessairement un et un seul des événements incompatibles $E_1, E_2, \ldots, E_n, \ldots$ dont les probabilités respectives sont $\frac{1}{2}$, $\frac{1}{2^2}, \ldots, \frac{1}{2^n}, \ldots$.

Appelons X_n la variable aléatoire qui est égale au nombre α_n si l'événement E_n se produit et à o dans le cas contraire; et supposons que α_n soit un nombre supérieur à l'unité quel que soit n.

Alors la suite $X_1, X_2, \ldots, X_n, \ldots$ converge « en probabilité » vers le nombre certain X égal à zéro. En effet, comme $X_n - X = 0$ ou α_n, la probabilité que $|X_n - X| \geqq \varepsilon$ est zéro si $\alpha_n < \varepsilon$ et sinon c'est la probabilité de E_n, soit $\frac{1}{2^n}$: elle tend dans tous les cas vers zéro quand n croît indéfiniment.

Prenons $r = 1$. L'écart moyen d'ordre 1, σ, de X avec l'unité est égal à 1. L'écart moyen d'ordre 1 de X_n avec l'unité est

$$\sigma_n = \frac{|\alpha_n - 1|}{2^n} + |0 - 1|\left(1 - \frac{1}{2^n}\right) = \frac{\alpha_n - 2}{2^n} + 1$$

Prenons, par exemple, $\alpha_n = 2$. Alors

$$\lim_{n \to \infty} \sigma_n = 1 = \sigma.$$

Mais nous pouvons prendre par exemple $\alpha_n = 2 + 2^n$, alors on aura

$$\lim_{n \to \infty} \sigma_n = 2 \neq \sigma.$$

Nous pouvons aussi prendre $\alpha_n = 2 + 4^n$, et alors

$$\lim_{n \to \infty} (\sigma_n = 1 + 2^n) \to \infty \neq \sigma$$

Nous pouvons enfin prendre $\alpha_n = 2 + [2 + (-1)^n] 2^n$. et alors $\sigma_n = 3 + (-1)^n =$ alternativement 2 et 4 de sorte que la suite des σ_n n'est pas convergente.

Application. — L'écart moyen $\lambda_n^{(r)}$ d'ordre r du nombre aléatoire X_n avec le nombre aléatoire X n'est autre que l'écart moyen d'ordre r de $X_n - X$ avec le nombre certain zéro

Par conséquent. d'après la généralisation connue de l'inégalité de Bienaymé, la probabilité $P_n(\varepsilon)$ pour que $|X_n - X| \geq \varepsilon$ est $\leq \left[\frac{\lambda_n^{(r)}}{\varepsilon}\right]^r$. Il en résulte que *si pour une valeur au moins du nombre positif r, la suite des écarts moyens $\lambda_n^{(r)}$ d'ordre r des X_n avec X tend vers zéro, le nombre aléatoire X_n converge « en probabilité » vers le nombre aléatoire X.*

La réciproque n'est pas exacte dans le cas le plus général Il suffit de reprendre l'exemple de la page précédente en prenant $\alpha_n = 2^{(2^n)}$ On voit alors que X_n converge probablement vers $X = 0$ et pourtant pour chaque valeur positive de r, l'écart moyen d'ordre r de X_n et X, soit $\lambda_n^{(r)} = 2^{\left(2^n - \frac{n}{r}\right)}$ croît indéfiniment avec n

On pourra d'ailleurs déterminer exactement le cas le plus général où la réciproque est exacte en appliquant les résultats obtenus plus haut page 185 après substitution de $X_n - X$ à X_n et de zéro à X. On voit alors que :

Si une suite de variables aléatoires X_n converge « en probabilité » vers la variable aléatoire X, la condition nécessaire et suffisante pour que l'écart moyen d'ordre r de X_n et X converge vers zéro quand n croît indéfiniment est que la convergence de l'intégrale $\int_0^A t^r d\,q_n(t)$ (vers la valeur moyenne de $|X_n - X|^r$ quand A croît indéfiniment) soit uniforme quand n varie (au moins à partir d'un certain rang n). [Ici $q_n(t)$ désigne la probabilité pour que $|X_n - X| < t$.]

La condition sera en particulier réalisée lorsque, au moins à partir d'un certain rang, les $|X_n|$ et $|X|$ (ou tout au moins les $|X_n - X|$) restent inférieures à un même nombre fixe. Mais elle le sera encore dans des cas plus étendus.

Un tel cas plus général se présente quand l'écart moyen d'ordre s de X_n et X reste borné à partir d'une valeur assez grande de n pour au moins une valeur de s supérieure à r. Car, en écrivant

$$(\lambda_n^{(s)})^s = \int_0^{+\infty} t^s \, dq_n(t),$$

on aura, pour toute valeur positive de B,

$$\int_B^{+\infty} t^r \, dq_n(t) \leq \frac{1}{B^{s-r}} \int_B^{+\infty} t^s \, dq_n(t) \leq \frac{1}{B^{s-r}} (\lambda_n^{(s)})^s \leq \frac{A_s}{B^{s-r}},$$

où A_s est indépendant de n.

Cela suffit pour assurer la convergence uniforme désirée.

Nous voyons en même temps que si X_n converge en probabilité vers X, trois cas peuvent se présenter en ce qui concerne l'écart moyen d'ordre r de X_n et de X : I, ou bien cet écart reste, pour chaque valeur de r, non borné quand n croît; II, ou bien il est borné pour une valeur s de r, non borné pour chaque valeur fixe de $r > s$ et il tend vers zéro pour $r < s$; III, ou bien, enfin, il tend vers zéro quel que soit r.

Autre forme de la condition. — La condition nécessaire et suffisante donnée quelques lignes plus haut peut d'ailleurs être exprimée sous une forme parfois plus commode dans le cas où le moment d'ordre r de $|X|$, soit $\mathfrak{M}|X|^r$, est fini. Il suffit pour cela d'utiliser une inégalité (160^c) que nous allons établir et qui exprime, en somme, que si les grandes valeurs de deux variables aléatoires Y et Z sont rares, il en est de même de leur somme (et de leur différence).

Partons d'abord de ce fait que si a et b sont deux nombres quelconques, il existe [1] au moins un nombre k, indépendant de a et de b tel que

$$(160^a) \qquad |a+b|^r \leq k_r \left\{ |a|^r + |b|^r \right\}$$

---

[1] Comme on a $|a+b|^r \leq \{|a|+|b|\}^r$, il suffit de prouver l'existence de k, quand a et b sont ≥ 0, et même puisque la relation a lieu quel que soit k, quand a et b sont tous deux nuls, il suffit de prouver que $\frac{(a+b)^r}{a^r+b^r}$ est borné quand a et b sont ≥ 0 et non tous deux nuls. Par suite de la symétrie, on peut supposer $0 \leq a \leq b$. En posant $t = \frac{a}{b}$, il suffit de prouver que $\frac{(1+t)^r}{1+t^r}$ a une borne pour $0 \leq t \leq 1$. Or,

Représentons maintenant en général par X_A un nombre égal à X quand $|X| \geqq A$ et à zéro quand $|X| < A$. La condition considérée ci-dessus revenait à dire que $\mathfrak{M}\{|X - X_n|^r\}_A$ converge vers zéro, uniformément quand n est assez grand ($n > N$) lorsque A croît indéfiniment. Or si $\mathfrak{M}|X|^r$ est fini, $\mathfrak{M}\{|X|^r\}_A$ converge aussi vers zéro avec $\frac{1}{A}$. L'inégalité (160^a) va être transformée de façon à en déduire une propriété analogue pour $\mathfrak{M}\{|X_n|^r\}_A$. Pour deux variables aléatoires Y et Z quelconques et $B > 0$ quelconque, il existe [1] un nombre L ne dépendant que de r et tel que

$$(160^b) \qquad \{|Y + Z|^r\}_{(2B)^r} \leqq L\{|Y|^r\}_{B^r} + L\{|Z|^r\}_{B^r}.$$

Ceci étant, on aura donc

$$(160^c) \qquad \mathfrak{M}\{|Y + Z|^r\}_{(2B)^r} \leqq L\mathfrak{M}\{|Y|^r\}_{B^r} + L\mathfrak{M}\{|Z|^r\}_{B^r},$$

en particulier

$$(160^d) \qquad \mathfrak{M}|Y + Z|^r \leqq L\mathfrak{M}|Y|^r + L\mathfrak{M}|Z|^r$$

Ceci nous montre d'abord, pour $Y = X$, $Z = a$, que si $\mathfrak{M}|X|^r$ est fini, il en est de même de $\mathfrak{M}|X + a|^r$ quelle que soit la quantité certaine a. En outre, on a

$$(160^e) \qquad \mathfrak{M}\{|X_n|^r\}_{(2B)^r} \leqq L\mathfrak{M}\{|X|^r\}_{B^r} + L\mathfrak{M}\{|X_n - X|^r\}_{B^r}.$$

---

cela résulte de ce que cette fonction est continue dans cet intervalle. On peut prendre pour k_r son maximum, on voit facilement que ce maximum est 2^{r-1} pour $r \geqq 1$ et 1 pour $r \leqq 1$

Quand on ne vise qu'à établir l'existence de k_r, on peut, plus simplement, observer avec M. Cramér, que si, par exemple, $|a| \geqq b$, on a

$$|a + b|^r \leqq (|b| + |b|)^r \leqq 2^r(|a|^r + |b|^r).$$

[1] En effet, prenons $C = 2B$ et $L = 2^r$, d'où $LB^r = C^r$,

I, ou bien $|Y| < B$ et $|Z| < B$ et alors $|Y + Z| < C$, l'inégalité (160^b) se réduit à $0 \leqq 0$;
II, ou bien $|Y| \geqq B$ et $|Z| \geqq B$, alors on a

$$\{|Y + Z|^r\}_{C^r} \leqq |Y + Z|^r \leqq k_r|Y|^r + k_r|Z|^r = k_r\{|Y|^r\}_{B^r} + k_r\{|Y|^r\}_{B^r}$$

et il suffit de prendre $L \geqq k_r$,

III, ou bien B est entre $|Y|$ et $|Z|$, par exemple $|Y| \leqq B < |Z|$, alors

$$\{|Y + Z|^r\}_{C^r} \leqq |Y + Z|^r \leqq (2|Z|)^r = 2^r\{|Y|^r\}_{B^r} + 2^r\{|Z|^r\}_{B^r}$$

et il suffit de prendre $L \geqq 2^r$. Comme on peut supposer $k_r \leqq 2^r$, on voit qu'on peut supposer $C^r = LB^r$ et $L = 2^r$.

En prenant d'abord $B = 0$, on voit que si $\mathfrak{M}|X_n - X|^r$ tend vers zéro avec $\frac{1}{n}$, il suffit que $\mathfrak{M}|X^r|$ soit fini pour que $\mathfrak{M}|X_n|^r$ soit aussi fini à partir d'un certain rang N et même ait à partir de ce rang une borne indépendante de n. En outre, dans les mêmes hypothèses, $\mathfrak{M}\{|X|^r\}_{B^r}$ sera aussi petit que l'on veut pour B assez grand, et par suite $\mathfrak{M}\{|X_n|^r\}_A$ va converger vers zéro quand $A \to \infty$ pour $n > N$. Il va même pour $n > N$ converger uniformément, car pour ε positif arbitraire, on pourra prendre B assez grand pour que

$$\mathfrak{M}\{|X|^r\}_{B^r} < \frac{\varepsilon}{2L},$$

puis N′ assez grand (N′ > N) pour que

$$\mathfrak{M}|X_n - X|^r < \frac{\varepsilon}{2L}$$

pour $n > N'$. Alors, on aura

$$\mathfrak{M}\{|X_n|^r\}_{(2B)^r} < \varepsilon$$

pour $n > N'$.

On pourra d'autre part. pour chaque valeur de $n > N$, assigner un nombre A_n, tel que $\mathfrak{M}\{|X_n|^r\}_A < \varepsilon$, pour $A > A_n$. Si maintenant on prend pour A le plus grand des nombres $A_N, A_{N+1}, \ldots, A_{N'}$ et $(2B)^r$, on aura

$$\mathfrak{M}\{|X_n|^r\}_A < \varepsilon$$

pour $n > N$, N, contrairement à N′, étant *indépendant* de ε. Inversement, si $\mathfrak{M}\{|X_n|^r\}_A$ converge vers zéro quand $A \to \infty$, uniformément à partir de n assez grand, et si $\mathfrak{M}\{|X_n|^r\}$ est fini, alors $\mathfrak{M}\{|X_n - X|^r\}_A$ converge vers zéro avec $\frac{1}{A}$ et cela, uniformément pour n assez grand. D'où le résultat suivant :

Quand X_n converge en probabilité vers X, alors, *si* $\mathfrak{M}|X|^r$ *est fini*, la condition nécessaire et suffisante pour que l'écart moyen d'ordre r de X_n et de X converge vers zéro est que l'intégrale représentant le moment d'ordre r de $|X_n|$ soit uniformément convergente à partir d'une certaine valeur de n.

Une condition suffisante pour qu'il en soit ainsi est que les $|X_n|$ soient bornés à partir d'un certain rang ou plus généralement qu'il existe un exposant $s > r$, tel qu'à partir d'un certain rang les moments d'ordre s des $|X_n|$ soient bornés dans leur ensemble.

Remarque. — Ce qui précède montre que l'écart moyen d'un ordre r quelconque, mais déterminé, de deux nombres aléatoires X, Y, jouit vis-à-vis de la « convergence en probabilité » de Y vers X de plusieurs des propriétés de la distance de deux points x, y relativement à la convergence du point y vers le point x.

Si X et Y sont « presque certainement » égaux, leur écart moyen d'ordre r, qu'on peut mettre sous la forme

$$\sqrt[r]{\int_0^{+\infty} t^r \, d\, q(t)},$$

où $q(t) = 1$ pour $t > 0$, est évidemment nul. Réciproquement, s'il est nul, $q(t)$ doit être égal à un pour $t > 0$ X et Y sont « presque toujours » égaux.

Si l'écart moyen d'ordre r de X et X_n converge vers zéro avec $\frac{1}{n}$, X_n converge « en probabilité » vers X.

Si X_n converge « en probabilité » vers X et si l'intégrale dont la limite permet de calculer cet écart moyen $\lambda_n^{(r)}$ a une convergence uniforme ou plus simplement si les X_n sont bornés, alors cet écart moyen tend vers zéro

Toutefois, nous voyons qu'il y a là une restriction qui nous empêche de déterminer s'il y a ou non convergence « en probabilité » de X_n vers X, moyennant la seule connaissance de leur écart moyen d'un certain ordre r ou même si l'on connaît leurs écarts moyens de tous les ordres.

C'est là un défaut grave de la notion d'écart moyen d'un ordre numérique donné. Nous allons donner maintenant une définition de la « distance » de deux variables aléatoires, qui échappe à ce défaut.

SECTION III . PREMIER ESPACE DE VARIABLES ALÉATOIRES.

Nous allons profiter une fois de plus de l'avantage qui consiste à avoir démontré une fois pour toutes, une série de propriétés d'une famille d'espaces abstraits : on pourra étendre ensuite, d'un coup et sans nouveau raisonnement, toutes ces propriétés à des espaces d'une nature déterminée si l'on peut prouver qu'ils appartiennent à cette famille.

Définition du nouvel espace. — Nous pouvons considérer toutes les variables aléatoires X qu'on peut définir dans une certaine catégorie d'épreuves comme autant de « points » d'un certain espace. On va étudier, dans ce volume, diverses définitions de la limite d'une suite de variables aléatoires : chacune d'elle complètera la définition d'un espace de variables aléatoires. Et les espaces ainsi définis ne seront pas identiques.

Ayant étudié dans ce qui précède une première définition de la limite, nous sommes en position de définir un premier espace de variables aleatoires. Et nous dirons qu'une suite de points $X_1, X_2, \ldots$ de cet espace converge vers le point X de cet espace quand la variable aléatoire X_n tend « en probabilité » vers la variable aléatoire X. Seulement, pour que l'unicité de la limite subsiste, nous sommes amenés à considérer, comme « *points* » *non distincts*, deux points représentés par deux variables aléatoires qui sont « presque certainement » égales ([1]).

Pour justifier cette façon de voir, nous devrons encore démontrer que si, quel que soit n, les nombres aléatoires X_n et Y_n sont « presque certainement » égaux : 1° la suite des X_n et celle des Y_n convergent « en probabilité » toutes deux ou aucune des deux. 2° dans le premier cas, leurs limites « en probabilité » sont « presque certainement » égales.

Supposons que X_n converge « en probabilité » vers X, alors ε, η étant deux nombres positifs donnés, il y a un entier N tel que la probabilité que $|X_n - X| \geqq \eta$ soit $< \varepsilon$ pour $n > N$. Comme il y a une probabilité nulle que $|X_n - Y_n| > 0$ et comme

$$X - Y_n = (X - X_n) + (X_n - Y_n)$$

alors, en vertu du lemme de la page 162, la probabilité que $|X - Y_n| > \eta$ est $< \varepsilon$ pour $n > N$. Donc, Y_n tend « en probabilité » vers X. Et si Y_n tend « en probabilité » vers Y, Y est, comme nous l'avons vu page 165, « presque toujours » égal à X.

Il serait utile de démontrer que l'espace des variables aléatoires est un espace (L) (E. A., p. 164) ([2]) c'est-à-dire ici que :

([1]) On rencontre ici une conception analogue à celle qui consiste à considérer deux fonctions mesurables comme non distinctes lorsqu'elles sont égales « presque partout ».

([2]) Dans la suite, je renverrai encore pour plusieurs définitions et théorèmes à

Si une suite de variables aléatoires X_1, X_2, ... *est formée de variables « presque certainement » égales à la variable aléatoire* X, *cette suite converge « en probabilité » vers* X;

Si une suite de variables aléatoires Y_1, Y_2, ... *converge « en probabilité » vers la variable aléatoire* Y, *il en est de même de toute suite extraite de la suite des* Y_n.

La démonstration directe de ces deux propriétés ne donnerait aucune peine, mais elle va résulter indirectement de la propriété suivante : *l'espace des variables aléatoires est « distanciable »* (E. A., p. 62). En fait, nous allons le « distancier ».

Distance de deux variables aléatoires. — Il s'agit de prouver *qu'on peut associer à tout couple* X, Y *de variables aléatoires, un nombre* $(X, Y) = (Y, X) > 0$, *nombre qu'on appellera « distance » de* X *et de* Y *et qui devra satisfaire aux conditions suivantes :*

Si X *et* Y *sont « presque certainement » égaux, leur « distance » est nulle et inversement,*

Si X_n *tend « en probabilité » vers* X, *leur « distance »* (X_n, X) *tend vers zéro et inversement,*

Si X, Y, Z *sont trois variables aléatoires quelconques, leurs « distances » vérifient « l'inégalité triangulaire »*

$$(X, Y) \leqq (X, Z) + (Z, Y).$$

Tout espace distanciable est un espace (L), mais nous avons montré (E. A., p. 162) que la réciproque n'est pas vraie et que les espaces distanciables jouissent de propriétés importantes qui n'appartiennent pas à tous les espaces (L), de sorte qu'il y a utilité à établir les dernières propositions en italiques.

Du moment qu'on aura défini une distance (X, Y), on pourra en donner une infinité d'autres définitions vérifiant les mêmes conditions, mais fournissant des valeurs numériquement différentes des précédentes. Il serait intéressant de chercher la plus simple. Mais cela n'a aucune importance au point de vue de l'utilisation des propriétés topologiques des espaces distanciables, l'existence d'une dis-

---

mon livre : *Les espaces abstraits* (Gauthier-Villars, 1928) en le désignant par l'abréviation E. A.

tance et non la forme de son expression, étant alors le seul fait qui importe. Nous nous contenterons donc de donner *une* définition de la distance, laissant ouverte la question d'en trouver une plus simple [1].

Généralisant notre première définition de « la distance » de deux fonctions mesurables, *nous appellerons* (Fréchet, 135, p. 35) *distance* (X, Y) *de deux variables aléatoires*, la borne inférieure quand ε varie, de la somme $(X, Y)_\varepsilon + \varepsilon$ où ε est un nombre positif arbitraire et où l'on a désigné par $(X, Y)_\varepsilon$ la probabilité que $|X - Y| \geqq \varepsilon$.

On a évidemment

$$(X, Y) = (Y, X) \geqq 0$$

1° Si X et Y sont « presque certainement » égaux, on a $(X, Y)_\varepsilon = 0$ et la borne inférieure (X, Y) de $(X, Y)_\varepsilon + \varepsilon$ est bien nulle. Pour démontrer la réciproque et les autres conditions, il nous sera utile de faire une remarque :

Si $(X, Y) < \delta$, *alors* $(X, Y)_\delta < \delta$. Car il y aura ε positif tel que $\varepsilon + (X, Y)_\varepsilon < \delta$; d'où $\varepsilon < \delta$ et $(X, Y)_\varepsilon < \delta$. Comme $\varepsilon < \delta$, la probabilité que $|X - Y| \geqq \delta$ est au plus égale à la probabilité que $|X - Y| \geqq \varepsilon$: on a bien

$$(X, Y)_\delta < \delta$$

2° Si X_n tend « en probabilité » vers X et si ε et η sont des nombres positifs arbitraires, il y a N tel que $(X_n, X)_\varepsilon < \eta$ pour $n > N$, d'où

$$(X_n, X) \leqq \varepsilon + (X_n, X)_\varepsilon < \varepsilon + \eta \qquad \text{pour } n > N$$

Comme ε, η sont arbitraires, on a bien

$$\lim_{n \to \infty} (X_n, X) = 0$$

Réciproquement, démontrons que si la distance (Y_n, Y) tend vers zéro avec $\frac{1}{n}$, Y_n tend « en probabilité » vers Y. Soient ε, η, ω trois nombres positifs arbitraires. Il existe un nombre N, tel que $(Y_n, Y) < \omega$ pour $n > N$. Alors comme on l'a vu $(Y_n, Y)_\omega < \omega$ pour $n > N$. Prenons $\omega < \varepsilon$ et $\omega < \eta$, alors

$$(Y_n, Y)_\eta \leqq (Y_n, Y)_\omega < \omega < \varepsilon;$$

---

[1] Depuis que ces lignes ont été écrites, nous avons songé à une autre expression de la distance qui peut avoir quelques avantages et qui sera indiquée plus loin, p. 213, où on la rencontre comme une conséquence naturelle de la « f-convergence »

la probabilité que $|Y_n - Y| \geqq \eta$ est donc $< \varepsilon$ pour $\eta > N$. Ainsi Y_n tend « en probabilité » vers Y.

3° On a

$$Y - Z = (Y - X) + (X - Z).$$

D'après le lemme de la page 162, on a donc, quel que soit le nombre positif ε,

$$(Y, Z)_{\varepsilon+\eta} \leqq (Y, X)_\varepsilon + (Z, X)_\eta,$$

d'où

$$\varepsilon + \eta + (Y, Z)_{\varepsilon+\eta} \leqq [\varepsilon + (Y, X)_\varepsilon] + [\eta + (Z, X)_\eta]$$

On peut prendre ε, η de sorte que les deux crochets soient aussi voisins qu'on voudra de (Y, X) et de (Z, X) respectivement. On peut donc prendre $\omega = \varepsilon + \eta$ de sorte que $\omega + (Y, Z)_\omega$ soit au plus égale à une quantité aussi voisine qu'on voudra de (Y, X) + (Z, X).

D'où résulte

$$(Y, Z) \leqq (Y, X) + (X, Z)$$

Si T est un autre nombre aléatoire, on aura de même

$$(X, Z) \leqq (X, T) + (T, Z),$$

d'où

$$(Y, Z) - (X, T) \leqq (Y, X) + (Z, T)$$

Par suite, en permutant Y avec T et Z avec X, on voit que

$$|(Y, Z) - (X, T)| \leqq (Y, X) + (Z, T)$$

En particulier, on en conclut que *la distance* (Y, Z) *de deux nombres aléatoires*, Y, Z *n'est pas altérée quand on remplace ceux-ci par des nombres aléatoires* X, T *qui leur sont respectivement « presque sûrement » égaux*. Car on a vu que, dans ce cas, le second membre est nul.

Des propriétés qui définissent la distance, on déduit immédiatement aussi que :

Si les termes d'une suite de variables aléatoires $X_1, X_2, \ldots$ *sont « presque certainement » égaux à la variable aléatoire* X, *cette suite converge « en probabilité » vers* X;

Si une suite de nombres aléatoires X_n *converge « en probabilité » vers un nombre aléatoire* X, *toute suite extraite de la première converge aussi « en probabilité » vers le même nombre aléatoire* X.

On en déduit aussi que si (X, X_n) et (X_n, Y_n) tendent vers zéro, il en est de même de (X, Y_n). Autrement dit : *si le nombre aléatoire* X_n *tend « en probabilité » vers le nombre aléatoire* X *et si la probabilité que* $|X_n - Y_n| \geqq \varepsilon$ *tend vers zéro pour toute valeur positive de* ε, *alors le nombre aléatoire* Y_n *tend aussi en « probabilité » vers* X.

Ces deux propositions établies directement par M. Cantelli, dans le cas où X est certain, sont ici, même dans le cas général où X est aléatoire, des conséquences immédiates de l'existence d'une « distance ».

Nous allons maintenant montrer que l'espace des variables aléatoires est complet (E. A., p. 74) quand on y choisit, pour convergence d'une suite d'éléments, la convergence « en probabilité ». Plus précisément, nous allons établir que, *dans cet espace, le critère de Cauchy est valable* quand on prend, pour distance de deux éléments, la distance définie page 194.

Troisième critère de convergence en probabilite. — Il s'agit de prouver que : *pour qu'une suite d'éléments* $X_1, X_2, \ldots, X_n, \ldots$ *de l'espace converge* (vers un certain élément de cet espace), *il faut et il suffit que la distance* (X_n, X_m) *soit infiniment petite avec* $\frac{1}{n} + \frac{1}{m}$.

Il n'est pas utile de faire une démonstration spéciale à l'espace des variables aléatoires pour prouver que la condition est nécessaire. Il suffit d'invoquer à cet effet les propriétés générales (p. 193) de la distance. Quant à la condition suffisante, il suffit d'invoquer le théorème de Slutsky de la page 169. En effet, soient ε et η deux nombres positifs arbitraires, appelons δ le plus petit de ces deux nombres. Il y a alors par hypothèse, un nombre positif tel que

$$(X_n, X_m) < \delta \qquad \text{pour } \frac{1}{n} + \frac{1}{m} < \rho$$

D'après la page 194, on aura alors

$$\text{Pr.} \left\{ |X_n - X_m| > \delta \right\} < \delta < \eta.$$

Or

$$\text{Pr} \left\{ |X_n - X_m| > \varepsilon \right\} < \text{Pr.} \left\{ |X_n - X_m| > \delta \right\}.$$

On a donc

$$\text{Pr.} \left\{ |X_n - X_m| > \varepsilon \right\} < \eta \qquad \text{pour } \frac{1}{n} + \frac{1}{m} < \rho$$

La condition suffisante de M. Slutsky étant ainsi remplie, la convergence en probabilité des X_n en résulte.

Conditions pour qu'un ensemble de variables aléatoires soit compact « en probabilité » — Nous avons depuis longtemps (E. A., p. 275) introduit la notion d'ensemble compact Un ensemble est dit compact lorsque de chacun des sous-ensembles infinis on peut extraire une suite convergente. Pour qu'un ensemble soit compact, il faut qu'il satisfasse à des conditions qui diffèrent suivant la nature des éléments de l'ensemble et suivant celle de la convergence envisagée Nous avons obtenu ces conditions dans des cas nombreux (E. A., p. 116-123) et en particulier dans le cas de la convergence « en mesure » des fonctions mesurables (¹). Le raisonnement employé dans ce dernier cas va pouvoir être ici suivi dans ses grandes lignes avec les modifications convenables

Remarque préparatoire. — Soit X une variable aléatoire Nous supposerons que cette variable ne peut prendre que des valeurs finies (bornées ou non) de sorte que, si $F(x)$ est sa fonction de probabilité totale,

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1$$

Soient maintenant ε et ω deux nombres arbitraires On pourra trouver deux nombres A et B tels que

$$F(A) < \frac{\varepsilon}{2}, \qquad F(B) > 1 - \frac{\varepsilon}{2} \qquad \text{et} \qquad A < B.$$

De sorte que la probabilité qu'on ait

$$A \leqq X < B$$

sera supérieure à $1 - \varepsilon$ Soit maintenant φ un entier tel que $\frac{B-A}{\varphi} < \omega$, et soit E_i l'événement consistant en ce que

$$A + (i-1)\frac{B-A}{\varphi} \leqq X < A + i\frac{B-A}{\varphi}$$

et e_X l'événement consistant en ce que

$$X < A \qquad \text{ou} \qquad X \geqq B.$$

On voit que l'événement e_X a une probabilité $< \varepsilon$ et que si l'on se place dans les cas où e_X n'a pas lieu, alors X est borné et il y a un nombre fini d'événements $E_1, E_2, \ldots, E_\varphi$ qui épuisent la certitude et pour chacun desquels

(¹) *Sur les ensembles compacts de fonctions mesurables* (*Fund. Math.*, t. 9, 1927, p. 25-32)

les valeurs de X restent comprises entre deux valeurs différant de moins de ω

Cas d'une suite qui converge en « probabilité ». — Si l'on fixe les quantités ε et ω, à chaque nombre aléatoire Y correspondront de même des nombres A, B, φ et des événements $e_Y, E_1, \ldots, E_\varphi$ en nombre fini. Nous allons montrer que si l'on prend successivement pour Y les termes X_n d'une suite qui converge « en probabilité », on peut supposer que A, B, $E_1, \ldots, E_\varphi$ soient les mêmes pour tous ces termes, e_Y dépendant seul de Y

Soit X la limite « en probabilité » de X_n; ε', ω' étant donnés, définissons d'abord comme plus haut pour X seul, mais en remplaçant ε, ω par ε', ω', les nombres A, B, φ et les événements $E_1, \ldots, E_\varphi$, e_X Soit e''_n l'événement consistant en ce que

$$|X - X_n| \geqq \omega'.$$

On sait qu'il y a un nombre N tel que, pour $n > N$, la probabilité de e''_n soit $< \varepsilon'$.

D'autre part, définissons pour X_n comme pour X deux nombres A_n, B_n tels que la probabilité que $A_n \leqq X_n < B_n$ soit $> 1 - \frac{\varepsilon'}{N+1}$. Puis prenons pour $A' + \omega'$ le plus petit des nombres A, $A_1, \ldots, A_N$ et pour $B' - \omega'$ le plus grand des nombres B, $B_1, \ldots, B_N$. Alors il y a une probabilité $> 1 - 2\varepsilon'$ que l'on ait simultanément

$$A' \leqq X < B' \quad \text{et à la fois,} \quad A' \leqq X_n < B' \quad \text{pour} \quad s = 1, 2, \ldots, N$$

Pour $n > N$, on aura les mêmes inégalités quand on aura à la fois

$$A \leqq X < B \quad \text{et} \quad |X - X_n| \leqq \omega',$$

et le concours de ces deux événements a une probabilité supérieure à $1 - 2\varepsilon'$.

Pour toute valeur entière de n, appelons e'_n l'événement consistant en ce que l'on n'a pas simultanément

$$A' \leqq X < B', \quad A' \leqq X_s < B \quad \text{pour} \quad s \leqq N.$$

Appelons enfin, lorsque $n > N$, e_n l'événement (e''_n ou e'_n) et posons $e_n = e'_n$ si $n \leqq N$. Alors pour chaque valeur de l'entier n, on a

$$A' \leqq X < B, \quad A' \leqq X_s < B \quad \text{pour} \quad s \leqq N \quad \text{et} \quad A' \leqq X_n < B',$$

sauf, peut-être, lorsque se produit un événement e'_n dont la probabilité est au plus égale à $4\varepsilon'$.

Enfin prenons φ tel que $\frac{B' - A'}{\varphi} < \omega'$, appelons pour $i \leqq \varphi$, E_i l'événement où a lieu

$$A' + (i-1)\frac{B' - A'}{\varphi} \leqq X < A' + i\frac{B' - A'}{\varphi} \tag{161}$$

et $E_i^{(n)}$ l'événement où a lieu

$$A' + (i-1)\frac{B' - A'}{\varphi} \leqq X_n < A' + i\frac{B' - A'}{\varphi}. \tag{162}$$

Soient enfin $E'_1, E'_2, \ldots, E'_{q'}$ les événements distincts en nombre fini du type

$$E_i E_j^{(1)} E_k^{(2)} \ldots E_m^{(N)}.$$

Quand $n \leqq N$, chacun des événements E'_s est conséquence de l'un des événements $E_j^{(n)}$, c'est-à-dire que X_n restera quand E'_s se produit entre deux bornes différant de moins de ω'.

Quand $n > N$, si l'un des événements E'_s a lieu sans qu'ait lieu e_n, c'est que l'un des événements E_i a lieu sans qu'ait lieu e_n. Autrement dit, on a les inégalités (161) et l'on a

$$X - \omega' < X_n < X + \omega',$$

d'où

$$A' + (i-1)\frac{B'-A'}{\varphi'} - \omega' \leqq X_n < A' + i\frac{B'-A'}{\varphi'} + \omega'$$

Autrement dit, X_n reste dans ces épreuves compris entre deux nombres différant de moins de $3\omega'$

Prenons maintenant $\varepsilon' = \frac{\varepsilon}{3}$ et $\omega' = \frac{\omega}{3}$. On voit que ε et ω étant deux nombres arbitrairement fixés, on a pu déterminer des nombres A', B', q et des événements en nombre fini $E'_1, \ldots, E'_{q'}$, et associer à chaque nombre aléatoire X_n un événement e_n de sorte que

1° pour chaque valeur de n la probabilité de e_n est inférieure à ε,

2° il y a deux bornes A', B', indépendantes de n, qui ne sont pas dépassées par X_n sauf peut-être quand e_n a lieu,

3° pour chacun des événements E'_s, l'oscillation de X_n reste inférieure à ω sauf peut-être quand e_n a lieu,

4° de plus, dans tous les cas, quand e_n n'a pas lieu, l'un au moins des $E'_1, \ldots, E'_{q'}$ se produit

Par analogie avec le cas des fonctions mesurables cité plus haut, nous exprimerons ces propriétés en disant que les nombres aléatoires X_n sont « *également* » « *presque* » *bornés* (propriété 1° et 2°) *et* « *également* » « *presque* » *régulièrement étalés* (propriété 1°, 3° et 4°) dans leur champ de variation

Cas d'un ensemble quelconque de variables aléatoires — Supposons maintenant que ces propriétés soient possédées par un ensemble F de variables aléatoires Y. Autrement dit, nous supposons qu'à tout couple de nombres positifs ε, ω correspondent trois nombres A, B, q, un nombre fini q d'événements $E_1, E_2, \ldots, E_q$ (qui épuisent la certitude quand $A \leqq Y < B$) et un ensemble d'événements e_Y associés respectivement aux divers nombres Y de F de sorte que

a, la probabilité de e_Y est inférieure à ε,

b, Y ne dépasse pas les bornes A, B sauf peut-être si e_Y se produit;

c, pour chacun des événements E_s l'oscillation de Y reste inférieure à ω tant que e_Y n'a pas lieu;

d, l'un au moins des E_s se produit quand l'un des e_Y n'a pas lieu.

La condition est suffisante. — Nous allons montrer que, dans ces conditions, de tout ensemble infini I de variables aléatoires distinctes appartenant à l'ensemble F, on peut tirer une suite qui est convergente « en probabilité », la limite « en probabilité » de cette suite pouvant d'ailleurs, *ou non*, faire partie de F.

En effet, considérons une suite infinie de variables aléatoires distinctes appartenant à I, soient $Y_1, Y_2, \ldots, Y_n, \ldots$.

Si $E_1 - e_{Y_n}$ a lieu, c'est-à-dire si E_1 a lieu sans que e_{Y_n} se produise, Y_n reste entre A et B et l'oscillation de Y_n reste inférieure à ω. Soit dans ces conditions M_n la borne supérieure des valeurs de Y_n quand $E_1 - e_{Y_n}$ a lieu. On aura $A \leqq M_n \leqq B$. Donc on peut extraire de la suite des Y_n, une suite telle que la suite correspondante extraite des M_n converge. Et en ne retenant de cette suite que les termes d'un rang assez grand, on pourra même supposer que dans cette suite S_1, les différences $|M_n - M_{n'}|$ sont toutes inférieures à ω. Si Y_n et $Y_{n'}$ appartiennent à cette suite, alors lorsque $E_1 - e_{Y_n} - e_{Y_{n'}}$ a lieu, $E_1 - e_{Y_n}$ et $E_1 - e_{Y_{n'}}$ ont lieu, par suite les bornes supérieures de Y_n et $Y_{n'}$ diffèrent de moins de ω et les oscillations de Y_n et de $Y_{n'}$ sont inférieures à ω. Donc $|Y_n - Y_{n'}| < 3\omega$ quand $E_1 - e_{Y_n} - e_{Y_{n'}}$ a lieu. Si, maintenant, on opère sur cette suite S_1 de la même façon que sur la suite des Y_n, en faisant jouer à E_2 le rôle de E_1, on pourra extraire de S_1 une suite S_2 telle que pour deux nombres Y_n, $Y_{n'}$ de S_2, on ait $|Y_n - Y_{n'}| < 3\omega$ lorsque $E_2 - e_{Y_n} - e_{Y_n}$ a lieu ou encore aussi $E_1 - e_{Y_n} - e_{Y_{n'}}$. Et ainsi de suite. Au bout de q opérations, on aura extrait une suite $\sigma = S_q$ de variables aléatoires $Z_1, Z_2, \ldots$, appartenant à I et telles que $|Z_n - Z_{n'}| < 3\omega$ lorsque $E_s - e_{Z_n} - e_{Z_{n'}}$ a lieu pour $s = 1$, ou $= 2$, ..., ou $= q$, c'est-à-dire dans tous les cas où e_{Z_n} ni $e_{Z_{n'}}$ n'ont lieu. Autrement dit, l'événement $|Z_n - Z_{n'}| \geqq 3\omega$ n'a lieu que si e_{Z_n} ou $e_{Z_{n'}}$ a lieu; sa probabilité est inférieure à 2ε. On a donc

$$(Z_n, Z_{n'}) \leqq 3\omega + (Z_n, Z_{n'})_{3\omega} < 3\omega + 2\varepsilon.$$

Prenons $\varepsilon = \omega = \frac{1}{5}$; on pourra extraire de I une suite infinie σ_1 telle que la « distance » de deux quelconques de ses termes soit constamment inférieure à 1, prenons $\varepsilon = \omega = \frac{1}{2} \times \frac{1}{5}$, on pourra extraire de σ_1 une suite σ_2 telle que la distance de deux quelconques de ses termes soit inférieure à $\frac{1}{2}$, etc. Soit enfin σ_0 la suite constituée du premier terme de σ_1, du second terme de σ_2, etc. Ce sera une suite infinie, extraite de I, formée de termes distincts et telle que la « distance » de deux de ses termes soit inférieure à $\frac{1}{n}$ à partir du rang n. C'est donc (p. 196) une suite convergente extraite de I.

La condition est nécessaire. — Je dis maintenant que si un ensemble G de nombres aléatoires X est tel que, de tout ensemble infini de nombres X distincts

extraits de G on puisse extraire une suite convergente « en probabilité », l'ensemble G satisfait aux conditions précédentes

Soit d'abord ε un nombre positif arbitraire on peut déterminer un nombre L tel que $\text{Pr.}[|X| \geqq L] < \varepsilon$ pour tout X de G En effet, dans le cas contraire, pour tout entier p il y aurait au moins un élément X_p de G, tel que $\text{Pr}\ [|X_p| \geqq p] > \varepsilon$ De la suite des X_p on pourrait tirer une suite convergente « en probabilité » Pour les valeurs de p correspondant à cette suite il existe en vertu du lemme de la page 177 un nombre K indépendant de p, tel que $\text{Pr}\ [|X_p| \geqq K]$ reste $> \varepsilon$, alors que $\text{Pr}\ [|X_p| \geqq p] > \varepsilon$, d'où contradiction pour $p > K$.

Ceci étant, donnons-nous encore arbitrairement un nombre positif ω Nous avons démontré (E. A., p. 75) que dans un espace (D) complet, si un ensemble G est compact, il n'existe qu'un nombre fini (ou nul) d'éléments de G dont les distances mutuelles soient toutes supérieures à ω Soient donc $Y_1, Y_2, \ldots, Y_r$, r éléments de G tels que tout élément de G soit à distance inférieure à ω de l'un au moins de ces r éléments

Le raisonnement qu'on a fait (p. 197-199) pour une suite convergente « en probabilité » s'appliquerait en particulier pour la suite $Y_1, Y_2, \ldots, Y_{r-1}, Y_r, Y_r, \ldots, Y_r, \ldots$ Donc il existe r événements $e_1, \ldots, e_r$, trois nombres A', B', q, et q événements, soient $E_1, E_2, \ldots, E_q$ tels que pour $h = 1, 2, \ldots, r$:

a, $\text{Pr}\ e_h < \varepsilon$;

b, si e_h a lieu, on a $(Y_h \leqq A'$ ou $Y_h \geqq B')$,

c, l'oscillation de Y_h dans E_s est $< \omega'$ quand e_h n'a pas lieu pour $s = 1, 2, \ldots, q$,

d, l'un, au moins, des E_s a lieu quand l'un des e_h n'a pas lieu

Si maintenant Y est un élément quelconque de G, il y a au moins un des termes Y_h de $Y_1, \ldots, Y_r$, tel que

$$(Y, Y_h) < \omega',$$

d'où (1°, p. 194)

$$(Y, Y_h)_{\omega'} > \omega'$$

ou

$$\text{Pr}\ [|Y - Y_h| \geqq \omega'] < \omega'.$$

Soient e'_Y l'événement $|Y - Y_h| \geqq \omega'$, e_h l'événement $(Y_h \leqq A'$ ou $Y_h \geqq B')$, e_Y l'événement e'_Y ou e_h. Alors si e_Y n'a pas lieu, on a

$$|Y - Y_h| < \omega' \qquad \text{et} \qquad A' < Y_h < B',$$

d'où

$$A' - \omega' < Y < B' + \omega'.$$

En posant $A = A' - \omega'$ et $B = B' + \omega'$, on voit que, pour tout élément Y de G, on a $A < Y < B$ sauf peut-être si e_Y a lieu. De plus, la probabilité de e_Y est inférieure ou égale à la somme des probabilités de e'_Y et e_h, soit à $\omega' + \varepsilon'$.

D'autre part, quel que soit l'entier $s \leqq q$, si E_s a lieu sans e_Y, l'oscillation de Y_h reste inférieure à ω' et $|Y - Y_h| < \omega'$, de sorte que l'oscillation de Y reste inférieure à $3\ \omega'$.

Soient maintenant ε et ω deux nombres positifs arbitraires. Si l'on prend $\varepsilon' = \frac{\varepsilon}{2}$ et ω' inférieur à $\frac{\omega}{3}$ et $\frac{\varepsilon}{2}$, la probabilité de l'événement e_Y sera $< \varepsilon$ l'oscillation de Y dans chaque $E_\delta - e_Y$ sera inférieure à ω; enfin l'un, au moins, des E_δ a lieu quand l'un des e_Y ne se réalise pas.

En résumé, *la condition nécessaire et suffisante pour qu'un ensemble* G *de variables aleatoires soit tel que, de tout ensemble infini d'elements de* G *on puisse extraire une suite convergeant « en probabilite », est que les variables de* G *soient « également » « presque » bornees et que ces variables soient « également » « presque » régulièrement étalées dans leur champ de variation*

SECTION IV : CONVERGENCES DE DIVERSES NATURES ET ESPACES CORRESPONDANTS.

Convergence au sens ordinaire. — La définition usuelle de la convergence nous aurait conduit à dire qu'une suite de variables aléatoires X_n *converge « au sens ordinaire »* vers la variable aléatoire X, si pour chacune des épreuves de la catégorie considérée, et pour tout nombre $\eta > 0$, $|X_n - X| < \eta$ à partir d'un certain rang N. Le nombre N dépendra en général du résultat de l'épreuve considérée, aussi bien que de η. Si le nombre N ne dépend que de η, on dira que X_n converge *uniformément* vers X « *au sens ordinaire* » Par exemple si X est une variable aléatoire quelconque et si l'on pose $X_n = X + \frac{1}{n}$, X_n est une variable aléatoire qui converge uniformément « au sens ordinaire » vers la variable aléatoire X.

Les deux définitions sont distinctes. Par exemple, si toutes les épreuves possibles sont en infinité dénombrable et si l'on peut les numéroter et les désigner par $e_1, e_2, \ldots$, prenons $X_n = X + Y_n$, en désignant par Y_n un nombre égal à $\frac{p}{n}$, p étant le rang de l'épreuve où il faut définir X_n. Alors X_n converge vers X au sens ordinaire, mais non uniformément.

Si X_n converge *uniformément* vers X « au sens ordinaire ». la probabilité que $|X_n - X| \geqq \eta$ est rigoureusement nulle quand n est suffisamment grand. Par conséquent X_n converge aussi vers X « en probabilité ». Si la convergence a lieu « au sens ordinaire » sans être uniforme, alors, pour chaque valeur de n, aussi petit que soit η, il peut y avoir des épreuves pour lesquelles $|X_n - X| \geqq \eta$ et la probabilité de cet événement n'est pas nécessairement nulle. Il n'est même pas évident qu'elle tende vers zéro.

Mais dans ce cas, la valeur aléatoire $(X_n - X)$ converge « au sens ordinaire » vers le nombre certain zéro.

Or, en généralisant un important théorème de M. Borel, M. Cantelli a démontré que dans ce cas $X_n - X$ converge aussi vers zéro « au sens du calcul des probabilités » ([1]).

En résumé : si une suite de variables aléatoires X_n converge « au sens ordinaire » vers la variable aléatoire X, la convergence a lieu aussi « en probabilité ». L'exemple du théorème de Bernoulli montre d'ailleurs que la réciproque n'est pas vraie.

On peut même donner *un exemple d'une variable aléatoire* X_n *qui converge « en probabilité » vers une variable aléatoire* X *sans converger « au sens ordinaire » pour quelque épreuve que ce soit.*

Soit, en effet, $f_n(x)$, une certaine fonction de x. Prenons pour X_n la valeur que prend $f_n(x)$ lorsqu'on donne à x une valeur numérique prise au hasard à l'intérieur de l'intervalle 0, 1, et dans des conditions telles que la probabilité de l'inégalité $\alpha \leqq x \leqq \beta$ soit égale à $\beta - \alpha$.

Considérons les nombres rationnels distincts

$$\frac{1}{2}, \quad \frac{1}{3}, \quad \frac{2}{3}, \quad \frac{1}{4}, \quad \frac{3}{4}, \quad \ldots, \quad \frac{1}{p}, \quad \frac{2}{p}, \quad \ldots, \quad \frac{p-1}{p}, \quad \ldots$$

et soit $\frac{p}{q}$ celui qui a le rang n. Prenons

$$f_n\left(\frac{p}{q}\right) = 1, \qquad f_n(0) = f_n\left(\frac{p-1}{q}\right) = f_n\left(\frac{p+1}{q}\right) = f_n(1) = 0,$$

et supposons que la courbe $y = f_n(x)$ soit la ligne polygonale dont les sommets viennent d'être définis. Alors X_n tend « en probabilité » vers le nombre certain $X = 0$. Car

$$\Pr\,[\,|X_n - X| \geqq \eta\,] \leqq \Pr\,[\,|X_n - X| > 0\,] = \frac{2}{q}$$

et il est clair que q croît indéfiniment avec n.

Pourtant X_n ne converge jamais vers X. Car, si $0 < \xi < 1$, il y a un nombre rationnel de dénominateur donné q, soit $\frac{p}{q}$, tel que

$$\frac{p - \frac{1}{2}}{q} \leqq \xi < \frac{p + \frac{1}{2}}{q}.$$

([1]) Ce même résultat est aussi obtenu par nous, page 227, d'une autre façon.

Si $\frac{p}{q}$ a le rang m, on aura $f_m(\xi) \geqq \frac{1}{2}$. Il est clair que si q croît indéfiniment, m aussi. Dès lors, $f_n(\xi) \geqq \frac{1}{2}$ pour une certaine infinité de valeurs m de n. Ainsi, pour chaque épreuve, $X_n \geqq \frac{1}{2}$ pour une infinité de valeurs m de n : X_n ne converge jamais au sens ordinaire vers X.

Faisons une remarque en passant. L'écart quadratique moyen μ_n de $X_n - X$ étant la racine carrée de la valeur moyenne de $(X_n - X)^2$, on a ici

$$\mu_n^2 = \int_0^1 [f_n(x)]^2 \, dx = 2\int_{\frac{p-1}{q}}^{\frac{p}{q}} [qx - p + 1]^2 \, dx$$

$$= \frac{2}{3q}\left[(qx - p + 1)^3\right]_{\frac{p-1}{q}}^{\frac{p}{q}} = \frac{2}{3q}.$$

Lorsque n croît indéfiniment, q aussi, donc μ_n tend vers zéro D'ailleurs, on tire de l'inégalité

$$\frac{1}{2}(X_n - X_{n'})^2 \leqq X_n^2 + X_{n'}^2,$$

la formule

$$\frac{1}{2}\overline{(X_n - X_{n'})^2} \leqq \mu_n^2 + \mu_{n'}^2.$$

Il en résulte que, pour la suite envisagée, $\overline{(X_n - X_{n'})^2}$ tend vers zéro quand le plus petit des nombres n et n' croît indéfiniment.

Quasi-équivalence asymptotique. — Nous avons vu (p. 170), que si la variable aléatoire X_n converge en probabilité vers la variable aléatoire X, la fonction des probabilités totales $F_n(x)$ de X_n converge vers celle, $F(x)$, de X, au moins en tout point de continuité de $F(x)$. Mais nous avons vu aussi (p. 174), que la réciproque n'est pas vraie.

On aurait donc une définition de la convergence plus générale que la convergence en probabilité, si l'on disait qu'on reconnaît cette nouvelle nature de convergence à la réalisation de la *seule* condition

$$F(x) = \lim_{n \to \infty} F_n(x)$$

[aux points de continuité de $F(x)$]. C'est là une idée assez naturelle contre laquelle nous avons cru devoir mettre en garde (Fréchet, 148).

Il est théoriquement légitime de définir arbitrairement la limite. Mais, il faut cependant que la définition ne heurte pas trop notre intuition. Or, l'exemple qui a été donné (p. 174) à l'occasion d'une réciproque, montre qu'en adoptant cette définition, on serait amené à considérer une suite X_n comme convergeant à ce nouveau sens vers X, même dans certains cas où X_n a pour chaque épreuve une limite au sens ordinaire et où cette limite est dans toute épreuve différente de X et *où l'on a même* $|X_n - X| = 1$ *dans toute épreuve et quel que soit n*.

Il n'est donc pas possible de retenir une telle définition de la limite. Cependant la propriété consistant dans l'égalité

$$F(x) = \lim_{n \to \infty} F_n(x)$$

correspond certainement à une analogie de *certaines* propriétés de X_n et X quand n est grand. Il serait donc légitime de rappeler cette analogie par un nom approprié, tel qu'équivalence asymptotique. D'ailleurs cette désignation ne nous paraît ni la meilleure, ni sans inconvénients, et nous préférerions l'expression quasi-équivalence asymptotique.

Convergence en moyenne quadratique. — A côté des modes de convergence définis ci-dessus, il en est un autre, celui de la convergence en moyenne quadratique, auquel on aurait pu être conduit aussi bien en restant dans le domaine propre des probabilités qu'en s'inspirant des conceptions modernes de l'analyse mathématique.

On sait toute l'importance en calcul des probabilités, de l'écart quadratique moyen d'une variable aléatoire avec sa moyenne ou avec un nombre certain quelconque. Nous avons même, plus haut, défini (p. 58, 61) l'écart quadratique moyen de deux variables aléatoires quelconques X et Y. On peut considérer celui-ci comme mesurant en quelque sorte la distance de ces deux variables aléatoires.

C'est là une idée d'autant plus naturelle que cet écart moyen, que nous pouvons aussi représenter par la notation (X, Y), jouit des propriétés suivantes :

I. $(X, Y) = (Y, X) \geq 0$.
II. Quand $X \equiv Y$, $(X, Y) = 0$.
III. Ou plus généralement : la condition nécessaire et suffisante

(*voir* p. 120) pour que $(X, Y) = 0$ est que X et Y soient équivalents (au sens de la p. 192), c'est-à-dire presque certainement égaux.

IV. $(X, Y) \leqq (X, Z) + (Z, Y)$.

Toutefois une difficulté se présente du fait que deux variables aléatoires n'ont pas toujours un écart quadratique moyen fini. Il faudra donc limiter cette conception de la distance à un espace abstrait dont chaque point abstrait est une variable aléatoire et choisi tel que deux quelconques des points de cet espace ait une distance.

Cet espace n'aura donc pas comme éléments toutes les variables aléatoires. Si, comme il est naturel, on veut que la variable constamment égale à zéro y figure, il faudra n'admettre que les X tels que $(X, 0)$ soit fini. Or $(X, 0) = \sqrt{\mathfrak{M} X^2}$. On peut appeler cette expression, la *norme* de X et la représenter par $\| X \|$. On aura donc à se limiter aux variables aléatoires ayant chacune une norme finie. D'ailleurs, en vertu de l'inégalité triangulaire, on a

$$(X, Y) \leqq (X, 0) + (0, Y) = \| X \| + \| Y \|.$$

Donc cette limitation est suffisante : deux variables aléatoires chacune de norme finie, ont un écart quadratique moyen fini.

On aurait aussi pu être conduit à la définition de cet espace par l'analogie avec l'espace des fonctions de carrés intégrables, où l'on appelle distance de deux telles fonctions $f(x)$, $g(x)$ définies sur l'intervalle a, b, l'expression

$$(f, g) = \sqrt{\frac{1}{b-a} \int_a^b [f(x) - g(x)]^2 \, dx}$$

On définit, dans cet espace fonctionnel, la convergence « en moyenne quadratique » qui joue un grand rôle dans la théorie des fonctions orthogonales, des équations intégrales, etc.

De même que dans cet espace fonctionnel, on dira qu'une suite de variables aléatoires X_n, converge *en moyenne quadratique* vers une variable aléatoire X, si la distance (X, X_n) tend vers zéro avec $\frac{1}{n}$, c'est-à-dire si

$$0 = \lim_{n \to \infty} \mathfrak{M}(X_n - X)^2 \quad [1].$$

---

[1] *Voir* Note C, Addition α, p. 293.

D'après l'inégalité de Bienaymé, on a

$$\Pr.\{|X_n - X| > \varepsilon\} < \frac{\mathfrak{M}(X_n - X)^2}{\varepsilon^2};$$

donc *si* X_n *converge en moyenne quadratique vers* X, X_n *converge aussi en probabilité vers* X.

Application. — Il sera souvent commode pour établir qu'une suite converge en probabilité de prouver qu'elle converge en moyenne quadratique.

Par exemple, soit v_n la moyenne arithmétique de n variables aléatoires indépendantes $Y_1, \ldots, Y_n$, soient ρ_n et μ_n les écarts quadratiques moyens de v_n et de Y_n. On a

$$\rho_n = \frac{\sqrt{\mu_1^2 + \ldots + \mu_n^2}}{n}.$$

Donc, si les écarts quadratiques moyens de $Y_1, \ldots, Y_n, \ldots$ restent bornés dans leur ensemble, la déviation $v_n - \overline{v_n}$ converge en moyenne quadratique vers zéro et par suite converge aussi en probabilité vers zéro. L'hypothèse sur les μ_n se trouve en particulier réalisée dans le cas important où les Y_n sont bornés dans leur ensemble. Elle est plus particulièrement réalisée quand les Y_n sont égaux à 0 ou 1, les v_n étant les fréquences d'un événement de probabilités déterminées $p_1, p_2, \ldots$ dans une suite d'épreuves indépendantes. On obtient alors le théorème de Poisson et comme cas particulier le théorème de Bernoulli.

Dans le cas simple où $Y_1, Y_2, \ldots, Y_n, \ldots$ sont les valeurs prises par une même variable aléatoire Y dans une suite d'épreuves indépendantes, le raisonnement fait plus haut établit la convergence en probabilité de $v_n = \frac{Y_1 + \ldots + Y_n}{n}$ vers $\overline{Y}$ pourvu que $\overline{Y}$ soit déterminé et que l'écart quadratique moyen de Y soit fini.

Ce résultat sera précisé plus loin (p. 255) en démontrant que dans ces conditions il y a même « convergence presque certaine ».

Ces deux résultats ont été étendus respectivement par M. Khintchine (3) et Kolmogoroff au cas où, $\overline{Y}$ restant déterminé, l'écart quadratique moyen de Y n'est pas supposé fini comme il sera prouvé page 257.

Réciproque. — Il est d'ailleurs facile de voir *a priori* qu'il ne peut cependant exister d'équivalence entre la convergence en probabilité et la convergence en moyenne quadratique. La première, en effet, a une signification déterminée pour une suite quelconque de variables aléatoires, définies sur la même catégorie d'épreuves, tandis que la seconde ne concerne que les variables aléatoires dont chacune a une norme finie.

On peut même former des exemples de suites convergeant en probabilité mais non en moyenne quadratique et où pourtant la norme de la limite et les normes des éléments de la suite, non seulement sont chacune finie, mais sont bornées dans leur ensemble. C'est le cas de l'exemple de la page 186, en y prenant $\alpha_n = \sqrt{2^n}$. On verra que dans ce cas la norme de X_n reste égale à 1, celle de X étant égale à zéro et que l'écart quadratique moyen de X_n et X, restant égal à 1, ne peut tendre vers zéro, bien que X_n converge en probabilité vers X.

Convergence en moyenne d'ordre r. — Sans être entré dans toutes les considérations précédentes, M. Cantelli (4, p. 18-21) avait nettement distingué dès 1916 la convergence en moyenne quadratique de la convergence en probabilité. Il avait en même temps fait observer que la conception de convergence en moyenne quadratique se généralise immédiatement sous la forme de convergence en moyenne d'ordre r. Il suffit de faire jouer à l'écart moyen d'ordre r le rôle de l'écart quadratique moyen. L'écart moyen d'ordre r étant une fonction croissante de r, on voit que s'il y a convergence d'ordre $r > 0$, il y a convergence pour tous les ordres positifs inférieurs à r.

L'espace considéré A_r sera celui des variables aléatoires X dont les normes d'ordre r sont chacunes finies, en appelant ainsi l'expression $\sqrt[r]{\mathfrak{M}|X|^r}$, et où une suite $X_1, X_2, \ldots$ a pour limite X quand X_n converge vers X en moyenne d'ordre r. Nous savons que cette convergence en moyenne d'ordre r entraîne la convergence en probabilité, mais non réciproquement.

Le théorème énoncé page 187 permet cependant de formuler la condition à laquelle doit satisfaire une suite de variables aléatoires X_n pour que la convergence en probabilité vers X y soit équivalente à la convergence en moyenne d'ordre r vers X : il faut et il suffit que les

intégrales représentant les valeurs moyennes des $|X_n - X|^r$ soient uniformément convergentes, à partir d'un rang n assez grand.

Quand on se restreint à l'espace A_r considéré ci-dessus, où chaque variable aléatoire a un moment d'ordre r fini, on peut simplifier la condition précédente, en utilisant la seconde forme de condition de la page 190.

D'ailleurs, comme $\mathfrak{M}|X_n|^r$ est fini pour chaque élément X_n de A_r, il suffit de se reporter à la démonstration de ce théorème pour voir qu'on pourra y prendre $N = 1$ et par suite supprimer la restriction : à partir d'un certain rang. Dès lors :

S*i chacune des variables aléatoires* X_n *a une norme d'ordre* r *finie, la condition nécessaire et suffisante pour que la convergence en probabilité des* X_n *vers une variable aléatoire dont la norme d'ordre* r *est finie, soit équivalente à la convergence en moyenne d'ordre* r, *est que les moments d'ordre* r *des* $|X_n|$ *soient des intégrales uniformément convergentes.* Pour obtenir ce corollaire des théorèmes des pages 187 et 208, il nous a suffi de nous inspirer d'une condition « d'égale sommabilité du carré d'une fonction » qui nous a été communiquée par M. Flamant (1) au sujet d'un probleme analogue dans le domaine des fonctions de carré sommable.

Ensembles compacts. — On peut, de ce qui précède, déduire la condition pour qu'un ensemble G d'éléments de l'espace A_r soit compact, c'est-à-dire tel que, de toute suite infinie d'eléments de G, on puisse extraire une suite qui converge en moyenne d'ordre r vers un élément de A_r.

Il faut d'abord (II) ce que nous appellerons la condition $e(H)$ que les moments d'ordre r des éléments de G soient des intégrales convergeant uniformément sur G. Sans quoi, il existerait un nombre $\varepsilon > 0$, tel que, quel que soit n, l'on n'ait pas $\mathfrak{M}\{|X|^r\}_A < \varepsilon$ pour tout X de G et tout $A > n$. Il existerait donc au moins un élément X_n de G, tel que $\mathfrak{M}\{|X_n|^r\}_n \geqq \varepsilon$. Alors, on pourrait extraire des X_n une suite de termes $Y_p = X_{n_p}$ telle que n_p croisse avec p, que

$$(162^a) \qquad \mathfrak{M}\{|Y_p|^r\}_{n_p} \geqq \varepsilon$$

et que la suite Y_p converge en moyenne d'ordre r vers une variable aléatoire Y de norme (d'ordre r) finie. Les résultats obtenus plus haut (p. 190) montrent qu'alors $\mathfrak{M}\{|Y_p|^r\}_A$ converge vers zéro avec $\frac{1}{A}$, et cela uniformément quand n varie. Donc le premier membre de (162^a) ne peut que converger vers zéro : on arrive bien à une contradiction.

Mais la convergence en moyenne d'ordre r entraînant la convergence en probabilité, d'après la page 208, on voit que G reste compact, quand on adopte pour convergence la convergence en probabilité.

Les conditions du théorème de la page 199 devront donc être réalisées. Les conditions a, b, c, d de cette page viennent donc s'ajouter à la condition e qui vient d'être obtenue pour constituer un ensemble de conditions nécessaires. Nous allons voir qu'elles sont suffisantes et même que la condition d entraînant b, celle-ci peut être supprimée. Finalement, il s'agit de démontrer que :

Pour qu'un ensemble G *de variables aléatoires de* A_r *soit compact, il faut et il suffit :*

I. *Que l'intégrale qui représente le moment d'ordre* r *de* $|X|$ *soit uniformément convergente quand* X *parcourt* G.

II. *Que les variables de* G *soient également presque régulièrement etalées.*

Nous venons de voir que l'ensemble des conditions I, II, est nécessaire. Montrons qu'il est suffisant. On le fera en montrant que la condition I entraîne que les variables de G sont également presque bornées (condition III). Si maintenant on considère une suite S d'éléments de G, il résultera des conditions II et III qu'on peut extraire de S une suite qui converge en probabilité, et d'après I cette suite convergera aussi en moyenne d'ordre r.

Pour déduire III de II, posons, en général,

$$Y = X - X_B, \quad \text{d'où} \quad |Y| \leqq B.$$

On a, d'après (160'),

$$\mathfrak{M}|X|^r \leqq L\mathfrak{M}|Y|^r + L\mathfrak{M}|X_B|^r \leqq LB^r + L\mathfrak{M}\{|X^r|\}_{B^r} \leqq LB^r + L\varepsilon = R,$$

en prenant B assez grand et indépendant de X d'après I. Donc les normes des X de G sont non seulement finies mais bornées dans leur ensemble. Dans ces conditions,

$$\text{Pr.}\{|X| > D\} \leqq \frac{\mathfrak{M}|X|^r}{D^r} < \frac{R}{D^r},$$

où R et D sont indépendants de X. Dès lors, on peut prendre D assez grand et indépendant de X, de sorte que $\frac{R}{D^r}$ soit inferieur à un nombre ε donné. Si l'on appelle f_X l'événement consistant en ce que $-D < X < D$, on voit qu'on aura Pr $f_X < \varepsilon$. Ainsi les X de G sont bien également presque bornés.

On observera que la condition I est nécessairement vérifiée (p. 188) quand les X sont uniformément bornés ou plus généralement quand, pour au moins une valeur de $s > r$, les moments d'ordre s des X sont bornés dans leur ensemble.

La méthode précédente peut être aussi employée dans l'étude des ensembles

compacts de fonctions ordinaires quand on y définit la convergence par la convergence en moyenne d'ordre r

D'ailleurs seuls les cas de $r = 1$ ou 2 sont pratiquement intéressants. Mais l'extension à r quelconque conduit tout naturellement à une généralisation nouvelle.

La « f-convergence » en moyenne.— On obtient cette généralisation, à savoir la convergence en moyenne relativement à une fonction $f(t)$, si, dans la définition de la convergence en moyenne quadratique, on substitue à l'écart quadratique moyen le « f-écart moyen » défini page 120.

L'emploi de l'inégalité (93) de la page 121, au lieu de l'inégalité de Bienaymé dans le raisonnement fait quelques lignes plus haut, montre que si une suite de variables aléatoires X_n « f-converge » en moyenne vers X, il y a aussi convergence « en probabilité »

L'exemple relatif à la moyenne quadratique montre que la réciproque n'est pas vraie pour toute « f-convergence ». Cependant M Slutsky a indiqué deux cas importants où la réciproque est vraie.

On est conduit à ces deux cas en cherchant à se débarrasser de l'objection *a priori* formulée plus haut. Pour qu'il y ait « f-convergence » quand il y a convergence en probabilité, il faut d'abord que celle-ci ait un sens et que par conséquent on soit assuré que les « f-écarts » considérés soient finis. Ils le sont sûrement, en particulier dans deux cas : celui où les valeurs numériques des X_n et de X sont bornées dans leur ensemble, celui où la fonction f est bornée. Ce sont les cas considérés par M. Slutsky (1). Dans l'un ou l'autre de ces deux cas, $f(|X_n - X)|$ reste inférieur à un nombre A indépendant de n. Or soient ε et η deux nombres positifs arbitraires. Si X_n converge en probabilité vers X, on a

$$\text{Pr.}\left\{|X_n - X| > \varepsilon\right\} < \eta$$

pour n assez grand ($n > N$). D'autre part le « f-écart moyen » e_n de X_n et X est déterminé par une relation de la forme

$$f(e_n) = \mathfrak{M} f(|X_n - X|) = \int_0^{+\infty} f(t)\, d\, q_n(t) = \int_0^{\varepsilon} + \int_{\varepsilon}^{+\infty} \leqq f(\varepsilon) + A\eta$$

pour $n > N$.

Donc le premier membre tend vers zéro avec $\frac{1}{n}$. Dès lors e_n doit

tendre vers zéro. Ainsi *dans l'un ou l'autre des deux cas, la convergence en probabilité entraîne la f-convergence en moyenne* et, en tenant compte aussi de la proposition directe, *elle lui est même équivalente*

Cas de réduction à la convergence en probabilité — C'est là une proposition très importante, elle montre en particulier qu'il suffit d'adjoindre aux conditions imposées antérieurement (p. 120, 121) à la fonction auxiliaire $f(x)$, la condition très large d'être bornée, pour que la f-convergence devienne *indépendante de la forme de f* et équivalente à la convergence en probabilité.

Définition d'un écart de deux nombres aléatoires. — Soit $\mathcal{F}$ la famille des fonctions $f(x)$ chacune continue, croissante, bornée, définie pour $x \geqq 0$ et nulle pour $x = 0$. Le « f-écart moyen » de deux variables aléatoires X, Y satisfera d'après les remarques faites plus haut (p. 120, 121 et 212) à son sujet aux conditions suivantes, que nous exprimerons en le désignant par la notation ((X, Y)). Si f appartient à $\mathcal{F}$:

1° ((X, Y)) = ((Y, X)) est un nombre fini $\geqq 0$, fini et déterminé pour chaque couple X, Y de variables aléatoires;

2° Si X, Y sont presque toujours égaux, ((X, Y)) = 0 et inversement;

3° Si X_n converge « en probabilité » vers X, ((X, X_n)) tend vers zéro et inversement.

Cet « f-écart moyen » joue donc le rôle de ce que nous avons appelé « l'écart » des deux éléments dans la Théorie des espaces abstraits (E. A., p. 214). La « f-convergence » reste. d'ailleurs inchangée, comme nous l'avons observé plus haut, quand on remplace f par une autre fonction de la famille $\mathcal{F}$, puisque cette « f-convergence » reste identique à la convergence en probabilité.

Cet écart moyen jouit, en outre, des propriétés suivantes qui lui sont communes avec les écarts moyens d'ordre r (p. 121), et qui sont précieuses en statistiques.

Si $|X - Y|$ est un nombre certain et si a est sa valeur, le « f-écart moyen » de X et de Y est égal à a.

Si $|X - Y|$ reste compris entre deux nombres certains b et c, le « f-écart-moyen » de X et de Y reste aussi compris entre b et c.

Si $|X - Y| \leqq |X_1 - Y_1|$ dans toute épreuve, « l'écart moyen » de X et de Y est au plus égal à celui de X_1 et de Y_1 lorqu'on les calcule tous deux relativement à la même fonction f.

Définition d'une nouvelle distance de deux variables aléatoires. — En renonçant aux deux premières de ces autres propriétés, on peut déduire de la définition du « f-écart moyen », la définition d'une nouvelle « distance » de deux variables aléatoires (Fréchet, 138).

Parmi les fonctions f de la famille $\mathcal{F}$, il en existe qui sont telles que, si b, c sont deux nombres non négatifs quelconques, on ait

$$f(b+c) \leqq f(b) + f(c)$$

Telles sont, par exemple, les fonctions $\frac{x}{1+x}$, $1 - e^{-x}$, thx; telles sont les fonctions $\frac{kx}{1+kx}$ pour toute valeur positive de k.

Soit $\mathcal{F}^\star$ la famille des fonctions de $\mathcal{F}$ qui jouissent de cette propriété supplémentaire.

Comme ce sont des fonctions croissantes, alors, si a est un nombre positif ou nul tel que $a \leqq b + c$, on aura pour ces fonctions

$$f(a) \leqq f(b) + f(c)$$

Or, quels que soient les nombres aléatoires X, Y, Z, on a toujours

$$|X - Y| \leqq |X - Z| + |Z - Y|,$$

par suite, si $f(x)$ appartient à $\mathcal{F}^\star$

$$f(|X - Z|) + f(|Z - Y|) - f(|X - Y|) \geqq 0$$

Le premier membre a, comme ses trois termes, une valeur moyenne finie bien déterminée (puisque f reste borné) qui sera nécessairement positive ou nulle. Si donc μ, ν, λ sont les trois « f-écarts moyens » correspondants, on a l'inégalité suivante qui constitue une propriété intéressante des écarts moyens relativement aux fonctions de $\mathcal{F}^\star$

$$f(\lambda) \leqq f(\mu) + f(\nu).$$

Mais nous pouvons aussi désigner par $[X, Y]$, l'expression

$$[X, Y] = \overline{f(|X - Y|)},$$

c'est-à-dire le f-moment de M. Slutsky, et alors on aura l'inégalité

triangulaire

$$[X, Y] \leqq [X, Z] + [Z, Y] \quad (163)$$

D'autre part, les propriétés 1°, 2°, 3° du « f-écart moyen » signalées ci-dessus (p. 212), appartiennent aussi au f-moment. Finalement, nous voyons que le « f-moment » de deux variables aléatoires X, Y, peut être pris comme « distance » de X et de Y dès que f appartient à la famille $\mathfrak{F}^*$. Par exemple, en prenant pour f la fonction th$x$, on aura l'expression entièrement explicite de la distance

$$[X, Y] = \overline{\text{th}\,|X - Y|}.$$

Nous avons préféré placer ici cette nouvelle expression de la distance pour bien marquer que les propriétés topologiques du premier espace des variables aléatoires, que nous avons développées pages 192-202, étaient basées sur la simple existence d'une « distance » compatible avec la convergence en probabilité et ne dépendaient pas de la forme plus ou moins simple de cette distance.

Il faut bien observer que les deux définitions données pour la distance, qui donnent des valeurs numériques différentes pour un même couple de variables aléatoires, sont des définitions relatives à la convergence en probabilité et qui ne conviendraient pas pour un autre mode de convergence.

Nouvelle distance de deux fonctions mesurables. — Il peut être intéressant de noter que si la première expression de la « distance » donnée plus haut nous a été suggérée par une expression analogue que nous avions obtenue antérieurement dans la théorie des fonctions, inversement l'expression actuellement fournie par le « f-moment » défini pour les besoins du Calcul des Probabilités nous conduit maintenant à une expression correspondante de la distance de deux fonctions mesurables.

Si $X = G(Z)$ et $Y = H(Z)$ et si Z est un nombre aléatoire assujetti à rester entre a et b avec répartition uniforme des probabilités, on aura

$$\overline{f(|X - Y|)} = \overline{f[|G(Z) - H(Z)|]} = \frac{1}{b-a}\int_a^b f[|G(t) - H(t)|]\,dt.$$

Et si $G(t)$, $H(t)$ sont des fonctions mesurables, si d'autre part f appartient à la famille $\mathfrak{F}^*$, $f[|G(t) - H(t)|]$ sera une fonction mesurable et bornée donc sommable et le dernier membre sera fini et bien déterminé.

C'est là une définition de *l'écart* de deux fonctions mesurables qui avait été déjà formulée directement par M. Paul Lévy en 1925 (2) sans passer par l'intermédiaire des probabilités.

Nous avons montré plus tard (Fréchet, 138) qu'on peut en déduire une

nouvelle définition de la *distance* de deux fonctions mesurables, celle qu'on obtient en posant

$$(G, H) = \int_a^b f[|G(t) - H(t)|]\,dt$$

et en assujettissant la fonction f à la condition supplémentaire

$$f(a) \leqq f(b) + f(c) \qquad \text{pour } a \leqq b + c.$$

C'est d'ailleurs déjà en 1921, que nous avions établi [*voir* note (1) de la page 161], la possibilité de définir (d'une première façon) la distance de deux fonctions mesurables (quand on convient de considérer comme suite convergente d'éléments de cet espace les suites de fonctions mesurables convergeant « *en mesure* »).

SECTION V CONVERGENCE PRESQUE CERTAINE.

Définition. — Étant données des variables aléatoires $X, X_1, X_2, \ldots, X_n, \ldots$, définies sur la même catégorie d'épreuves, le résultat de chaque épreuve assigne à ces variables une suite de valeurs numériques bien déterminées, pour lesquelles il peut y avoir, ou non, convergence de X_n vers X. La convergence de ces valeurs de X_n vers cette valeur de X est donc un événement fortuit qui, pour chaque épreuve, a lieu, ou non, et dépend — en général — du résultat de cette épreuve. Supposons que cet événement ait une probabilité déterminée. [Nous laissons ici de côté la question de savoir s'il y a là une hypothèse nouvelle. Cette question sera résolue (p. 245-246) par la négative] Nous pourrons alors introduire une nouvelle notion :

Étant données deux variables aléatoires X_n et X définies sur la même catégorie d'épreuves, nous dirons que X_n *converge « presque certainement » vers* X si la probabilité de la convergence de X_n vers X est égale à l'unité.

Théorème de M. Borel. — M. Borel paraît être le premier à avoir signalé, au moins dans un cas particulier, cette nouvelle forme de convergence que nous considérons maintenant d'une manière générale.

Dans le Mémoire fondamental de 1909, où il pose les bases de la théorie des probabilités dénombrables (Borel, 1), il écrit à la page 14, § 11 :

« Nous appellerons *fréquence* d'un chiffre décimal jusqu'au rang n,

le quotient par n du nombre de fois que ce chiffre figure dans les n premieres décimales; si la fréquence ainsi définie tend vers une limite lorsque n augmente indéfiniment, on dira que la *fréquence totale* existe et que sa valeur est égale à cette limite.

» Ces définitions posées, on a l'énoncé suivant :

» La probabilité *pour que la fréquence totale d'un chiffre déterminé existe et soit égale* à $\frac{1}{10}$ a pour valeur l'unité. »

Cet énoncé est établi dans l'hypothèse que, pour chaque chiffre déterminé et chaque rang déterminé, il y a une probabilité $p = \frac{1}{10}$ pour qu'à ce rang se présente ce chiffre. On voit donc que l'énoncé de M. Borel peut aussi s'exprimer ainsi :

« La fréquence d'un chiffre décimal jusqu'au rang n converge *presque certainement* quand n croît vers la probabilité constante qu'a ce chiffre de se présenter à un rang déterminé (quelconque). »

M. Borel avait d'ailleurs envisagé d'abord le cas du système de numération binaire. Supposons qu'on effectue une série d'épreuves où l'on choisit au hasard l'un des chiffres 0 ou 1; soit $f_n = \frac{r_n}{n}$, la fréquence dans les n premières épreuves de l'apparition du chiffre 1. Alors, quand n croît, f_n converge « presque certainement » vers $\frac{1}{2}$, c'est-à-dire vers la probabilité du chiffre 1 à un rang déterminé.

Le raisonnement de M. Borel [qui figure aussi dans ce Traité (II, 1, p. 30-33)] est, comme on peut s'en assurer, indépendant [1] de la valeur particulière $\frac{1}{2}$ qu'en vue de son application arithmétique on a attribué à la probabilité du choix du chiffre 1. Nous pouvons donc dire :

« Si, en choisissant au hasard à chaque épreuve l'un des chiffres 0 ou 1, il y a une probabilité constante p de choisir le chiffre 1, alors la fréquence f_n de ce chiffre dans les n premières épreuves converge presque certainement vers p. »

M. Borel n'a pas fait observer explicitement une conséquence de ce résultat qui est absolument immédiate, mais extrêmement impor-

---

[1] Bien entendu, en modifiant en conséquence les formules employées.

tante. C'est celle qu'on obtient en interprétant p comme la probabilité supposée constante d'un événement fortuit de nature quelconque autre que le choix d'un nombre. Rien ne sera changé aux raisonnements. Mais le raisonnement obtenu par ce simple changement de mots apporte au théorème de Bernoulli un complément d'importance capitale. Le dernier théorème ainsi complété prend la forme suivante :

La fréquence $f_n = \frac{r_n}{n}$, au cours de n épreuves, d'un événement fortuit E de probabilité constante p, converge « presque certainement » vers p quand le nombre n des épreuves croît indéfiniment.

C'est à M. Cantelli (5) que revient le mérite d'avoir plus tard démontré le premier un théorème plus général (*voir* p. 239), exprimant ce que M. Khintchine a plus tard appelé la « loi forte » des grands nombres. L'application de ce théorème au cas particulier de Bernoulli a permis à M. Cantelli d'énoncer le premier explicitement [1] un complément au théorème de Bernoulli qui est substantiellement équivalent, comme nous le verrons, au résultat précédent. La démonstration de M. Cantelli est simple, complète et rigoureuse; nous la rappellerons plus loin (p. 239). Toutefois, la démonstration de M. Borel offre l'avantage de se prêter peut-être mieux à des précisions sur l'évaluation de l'erreur commise en remplaçant la probabilité par la fréquence. Et comme elle est la première en date, nous l'exposerons la première.

Démonstration analytique. — La démonstration de M. Borel est excessivement brève. Elle omet plusieurs raisonnements intermédiaires et admet certains résultats sans démonstration. Vue l'importance du résultat, nous allons ici la rappeler et la compléter.

Elle comporte implicitement ou explicitement quatre parties.

I. Soit P la probabilité pour que f_n converge vers p; soit ϖ la probabilité pour que soit réalisée une infinité des événements représentés respectivement par

$$|f_1 - p| \geqq \eta_1, \quad \ldots, \quad |f_n - p| \geqq \eta_n, \quad \ldots$$

---

[1] *Voir* toutefois, page 226, l'interprétation donnée antérieurement par M. Hausdorff au théorème de M. Borel.

Si $\lim_{n \to \infty} \eta_n = 0$, le contraire du second événement entraîne le premier, on en conclut qu'on a $P \geqq 1 - \varpi$.

Alors, pour démontrer que $P = 1$, il suffit de montrer qu'on peut choisir une suite de nombres certains η_n tendant vers zéro avec $\frac{1}{n}$ et pour laquelle $\varpi = 0$.

II. Si Q_n est la probabilité d'un événement E_n, et si ϖ' est la probabilité pour que soit réalisée une infinité des événements E_n, alors il suffit que ΣQ_n converge pour que ϖ' soit nul. C'est ce que M. Borel a démontré, quand les événements E_n sont indépendants, qui a été étendu par M. Cantelli [5] au cas des événements dépendants ou non et qui a été établi plus haut (p. 26).

Nous sommes maintenant conduit à prendre pour E_n l'événement

$$|f_n - p| \geqq \eta_n$$

et à choisir les η_n, de sorte que ΣQ_n converge.

III. Posons

$$\eta_n = t_n \sqrt{\frac{2pq}{n}} \qquad (t_n > 0)$$

et

$$q_n = \frac{2}{\sqrt{\pi}} \int_{t_n}^{+\infty} e^{-\lambda^2} d\lambda$$

le rapport $\frac{Q_n}{q_n}$ a une borne supérieure finie quand n varie quand on prend les t_n convenablement. De sorte que pour établir la convergence de ΣQ_n, il suffit de démontrer que Σq_n converge dans les mêmes hypotheses sur la variation de t_n avec n. Ceci, en supposant $\lim_{n \to \infty} \eta_n = 0$, ce qui nécessite déjà qu'on ait choisi les t_n, de sorte que

$$\lim_{n \to \infty} \frac{t_n}{\sqrt{n}} = 0$$

IV. Pour que Σq_n converge, il faut d'abord que q_n tende vers zéro, donc que t_n croisse indéfiniment.

D'après les propriétés de l'intégrale q_n, les deux conditions imposées à t_n, et qui peuvent être réalisées en prenant, par exemple, $t_n = \mathcal{L} n$, suffisent pour entraîner la convergence de Σq_n.

Le théorème est alors établi.

Il reste maintenant à prouver deux assertions intermédiaires.

En ce qui concerne le dernier point IV, on sait que q_n est asymptotiquement équivalent à $\frac{1}{t_n\sqrt{\pi}}e^{-t_n^2}$. On vérifie même (¹) qu'on a pour $t > 0$

$$(164) \qquad \frac{2}{\sqrt{\pi}}\int_t^{+\infty} e^{-\lambda^2}\,d\lambda = \frac{1}{t\sqrt{\pi}}e^{-t^2}\left[1 - \frac{\theta(t)}{2t^2}\right] \qquad \text{où } 0 < \theta(t) < 1$$

Il suffit donc de prouver que $\sum_n \frac{e^{-t_n^2}}{t_n}$ converge, en prenant, par exemple, $t_n = \mathcal{L}n$. Or, cela résulte de ce que, dans ce cas,

$$\frac{e^{-t_n^2}}{t_n} = \frac{1}{n^{\mathcal{L}n}\,\mathcal{L}n} < \frac{1}{n^2},$$

pour n assez grand.

Il ne reste plus à élucider que le second point.

On sait que le rapport de

$$Q_n(\lambda) = \Pr\left[|f_n - p| > \lambda\sqrt{\frac{2pq}{n}}\right]$$

à

$$2\,K(\lambda) = \frac{2}{\sqrt{\pi}}\int_\lambda^{+\infty} e^{-t^2}\,dt$$

(¹) Posons

$$\lambda = t + u,$$

on aura pour l'intégrale $K = \frac{1}{\sqrt{\pi}}\int_t^{+\infty} e^{-\lambda^2}\,d\lambda$,

$$K = \frac{e^{-t^2}}{\sqrt{\pi}}\int_0^{+\infty} e^{-u^2-2ut}\,du$$

Posons

$$2ut = v;$$

on aura

$$2K = \frac{e^{-t^2}}{t\sqrt{\pi}}\int_0^{+\infty} e^{-\frac{v^2}{4t^2}}e^{-v}\,dv = \frac{e^{-t^2}}{t\sqrt{\pi}}\int_0^{+\infty}\left(1 - \frac{\theta_1 v^2}{4t^2}\right)e^{-v}\,dv$$

$$= \frac{e^{-t^2}}{t\sqrt{\pi}}\left\{1 - \frac{\theta}{4t^2}\int_0^{+\infty} v^2 e^{-v}\,dv\right\} \quad \text{avec} \quad 0 < \left\{\begin{matrix}\theta_1\\ \theta\end{matrix}\right\} < 1.$$

D'où

$$2K = \frac{e^{-t^2}}{t\sqrt{\pi}}\left\{1 - \frac{\theta}{2t^2}\right\} \qquad \text{avec} \qquad 0 < \theta < 1.$$

tend vers l'unité quand λ *restant fixe*, n croît. Il n'en résulte pas d'une façon évidente que ce rapport reste inférieur à un nombre fixe quand, en même temps que n, λ croît indéfiniment. Mais il en est pourtant ainsi quand λ croît avec n de façon convenable.

La condition $|f_n - p| \geqq \eta_n$ peut s'écrire,

$$|r_n - np| \geqq t_n \sqrt{2npq}$$

et l'on a

$$Q_n = [\varpi_0^{(n)} + \varpi_1^{(n)} + \ldots + \varpi_\rho^{(n)}] + [\varpi_r^{(n)} + \varpi_{r+1}^{(n)} + \ldots + \varpi_n^{(n)}] = Q''_n + Q'_n,$$

en appelant ρ et r des entiers convenables, r, par exemple, étant le plus petit entier vérifiant

$$r - np \geqq t_n \sqrt{2npq} \tag{165}$$

Il suffira alors de montrer que $\frac{Q'_n}{K(t_n)}$ et $\frac{Q''_n}{K(-t_n)}$ restent inférieurs à des nombres indépendants de n quand les t_n sont choisis convenablement de façon à vérifier les conditions simultanées

$$t_n \to +\infty, \qquad \frac{t_n}{\sqrt{n}} \to 0$$

Prenons, par exemple, le premier rapport $\frac{Q'_n}{K(t_n)}$. Pour borner Q'_n, observons que

$$\frac{\varpi_{u+1}^{(n)}}{\varpi_u^{(n)}} = \frac{n-u}{u+1}\,\frac{p}{1-p} \leqq \frac{n-r}{r+1}\,\frac{p}{1-p} = \lambda < 1$$

pour $u \geqq r$ et $r - np \geqq -q$.

La dernière condition est vérifiée en vertu de (165) où t_n est supposé positif. On a donc

$$Q'_n \leqq \varpi_r^{(n)}[1 + \lambda + \lambda^2 + \ldots] = \frac{\varpi_r^{(n)}}{1-\lambda};$$

d'où

$$\frac{Q'_n}{K(t_n)} \leqq \frac{\varpi_r^{(n)}}{K(t_n)}\,\frac{1}{1-\lambda}.$$

Or, on a vu (p. 219), que

$$K(t_n) = \frac{1}{2t_n\sqrt{\pi}} e^{-t_n^2}[1 - \theta_n],$$

où θ tend vers zéro avec $\frac{1}{t_n}$. D'autre part,

$$1-k = \frac{r-np+1-p}{(r+1)(1-p)} = \frac{\frac{r}{n} - p + \frac{1-p}{n}}{\left(\frac{r}{n}+\frac{1}{n}\right)(1-p)}$$

et

$$r - np = t_n\sqrt{2npq} + z_n \qquad \text{avec } 0 \leqq z_n < 1$$

Alors

$$\frac{r}{n} - p = \frac{t_n}{\sqrt{n}}\sqrt{2pq} + \frac{z_n}{n},$$

par suite, $\frac{r}{n}$ tend vers p, et le dénominateur de $1-k$ tend vers pq. Le numérateur est

$$\frac{t_n\sqrt{2pq} + [z_n + 1 - p]\frac{1}{\sqrt{n}}}{\sqrt{n}} \sim \frac{t_n}{\sqrt{n}}\sqrt{2pq}$$

On a donc

$$1-k = \frac{t_n}{\sqrt{n}}\sqrt{\frac{2}{pq}}(1+U_n),$$

où U_n tend vers zéro avec $\frac{1}{n}$. Dès lors,

$$\frac{Q'_n}{K(t_n)} \leqq \left[\frac{\varpi_r^{(n)}}{\frac{e^{-t_n^2}}{\sqrt{2\pi npq}}}\right]\frac{1+U_n}{1-\theta_n}.$$

Or, on a vu (p. 95), que si la quantité $x = \frac{(r-np)}{\sqrt{2npq}}$ est telle que $\frac{x}{\sqrt[6]{n}}$ tend vers zéro, on a

$$\omega_r^{(n)} = \frac{e^{-x^2}}{\sqrt{2\pi npq}}[1+h_r^{(n)}],$$

où $h_r^{(n)}$ tend vers zéro avec $\frac{1}{n}$.

Si l'on suppose que t_n tend vers l'infini, de sorte que $\frac{t_n}{\sqrt[6]{n}}$ converge vers zéro, il en sera de même pour $\frac{x}{\sqrt[6]{n}}$ puisque

$$0 \leqq x - t_n < \frac{1}{\sqrt{2npq}},$$

et cela suffira aussi pour que $\frac{t_n}{\sqrt{n}}$ tende vers zéro. Enfin, ces deux conditions plus précises pour t_n sont encore vérifiées dans le cas où $t_n = l^2 n$.

Ainsi,

$$\frac{Q'_n}{K(t_n)} \leqq e^{t_n^2 - x^2}[1 + h_1^{(n)}]\frac{1 + U_n}{1 - \theta_n};$$

et, puisque $x \geqq t_n$,

$$\frac{Q'_n}{K(t_n)} \leqq [1 + h_1^{(n)}]\frac{1 + U_n}{1 - \theta_n}.$$

On voit que les $\frac{Q'_n}{K(t_n)}$ ont bien une borne supérieure finie $\Big($et même, on peut préciser que la plus grande des limites de $\frac{Q'_n}{K(t_n)}$ est $\leqq 1\Big)$.

Voir aussi Note C : Addition (β), p. 295.

Remarque. — Le théorème de M. Borel étant maintenant démontré, on peut compléter le dernier point et en même temps les résultats des pages 97-104. Considérons le rapport

$$\frac{Q'_n(\lambda)}{K(\lambda)} = \frac{\Pr\{r - np \geqq \lambda\sqrt{2npq}\}}{\frac{1}{\sqrt{\pi}}\int_{\lambda}^{+\infty} e^{-u^2}\,du}.$$

On peut écrire, en posant $\lambda' > \lambda$,

$$\frac{Q'_n(\lambda)}{K(\lambda)} = \frac{P_n(\lambda, \lambda') + Q'_n(\lambda')}{J(\lambda, \lambda') + K(\lambda')}.$$

Supposons que $\frac{\lambda}{\sqrt[6]{n}}$ converge uniformément vers zéro et prenons par exemple $\lambda' - \lambda = \sqrt[7]{n}$. Alors $\frac{\lambda'}{\sqrt[6]{n}}$ va aussi converger uniformément vers zéro. Or il existe un nombre S tel que, dans ces conditions, on ait $Q'_n(\lambda') < SK(\lambda')$ comme on vient de le démontrer. De plus il est clair que $\frac{1}{(\lambda' - \lambda)\sqrt{n}}$ converge uniformément vers zéro. D'après la page 104, il en résulte que si l'on pose

$$\frac{P_n(\lambda, \lambda')}{J(\lambda, \lambda')} = 1 + \gamma(n)$$

$\gamma(n)$ converge uniformément vers zéro. En posant $Q'_n(\lambda') = \theta_n SK(\lambda')$, on a

$$0 < \theta_n < 1$$

et

$$\frac{Q'_n(\lambda)}{K(\lambda)} = \frac{[1+\gamma(n)]J(\lambda, \lambda') + \theta_n S K(\lambda')}{J(\lambda, \lambda') + K(\lambda')};$$

d'où

$$\frac{Q'_n(\lambda)}{K(\lambda)} = \frac{1+\gamma(n)+\theta_n S \dfrac{K(\lambda')}{J(\lambda, \lambda')}}{1+\dfrac{K(\lambda')}{J(\lambda, \lambda')}}.$$

Or

$$J(\lambda, \lambda') > \frac{(\lambda'-\lambda)e^{-\lambda'^2}}{\sqrt{\pi}}, \qquad K(\lambda') < \frac{1}{\lambda'\sqrt{\pi}}e^{-\lambda'^2},$$

d'où

$$\frac{K(\lambda')}{J(\lambda, \lambda')} < \frac{1}{\lambda'(\lambda'-\lambda)} < \frac{1}{\sqrt[3]{n^2}}.$$

Donc $\frac{K(\lambda')}{J(\lambda, \lambda')}$ converge uniformément vers zéro. Finalement *si* $\frac{\lambda}{\sqrt[6]{n}}$ *converge uniformément vers zéro avec* $\frac{1}{n}$, on a l'équivalence asymptotique

$$K(\lambda) \sim Q'_n(\lambda)$$

ou plus précisément : *en posant*

$$(166) \qquad \frac{\Pr\{r-np \geq \lambda\sqrt{2npq}\}}{\dfrac{1}{\sqrt{\pi}}\displaystyle\int_{\lambda}^{+\infty} e^{-u^2}\,du} = 1+\varepsilon_n$$

ε_n *converge uniformément vers zéro*. C'est le résultat annoncé page 105.

La démonstration précédente suppose $\lambda > 0$. Mais si $\lambda < 0$, le dénominateur de l'expression ci-dessus reste $> \frac{1}{2}$ et le résultat est une conséquence de la page 101. Enfin, si le signe de λ n'est pas préfixé et peut varier avec n, le résultat est une conséquence des deux résultats correspondant à $\lambda > 0$ et à $\lambda < 0$.

Démonstration géométrique. — Dans le même mémoire de Palerme, M. Borel fait l'observation suivante : « L'ensemble des deux hypothèses se justifie d'ailleurs aisément lorsqu'on se place, non au point de vue logique, mais au point de vue géométrique ; elles sont, en effet, équivalentes à la suivante : le nombre décimal étant représenté par un point du segment 0 — 1, *la probabilité, pour qu'il se trouve sur un segment partiel est égale à la longueur de ce segment*. On pourrait

interpréter et vérifier à ce point de vue les résultats que nous allons obtenir ; je n'y insisterais pas, afin de laisser entièrement de côté pour le moment la théorie des probabilités continues, qui se rattache, comme je l'ai montré ailleurs, à la théorie des ensembles. »

Dans ce passage, M. Borel indique la possibilité d'une seconde démonstration, géométrique, cette fois. M. Mirimanoff qui a attiré notre attention sur ce passage nous a fait observer que dans l'édition de 1914 de son livre (I, p. 419), M. Hausdorff a explicité une telle démonstration pour le théorème qu'il appelle théorème de Borel et qu'il énonce à peu près ainsi qu'il suit :

L'ensemble E des points x du segment (0, 1) pour lesquels la fréquence totale du chiffre 1, dans le développement de x dans le système binaire, est déterminée et égale à $\frac{1}{2}$ a pour mesure l'unité (¹).

Voici la démonstration de M. Hausdorff

Comme l'ensemble des nombres rationnels est de mesure nulle, il suffit de démontrer que la mesure de I est égale à 1, en appelant I l'ensemble des nombres irrationnels appartenant à E. Ceci aura l'avantage d'éliminer les nombres x qui ont deux représentations binaires.

L'ensemble des x ($0 < x < 1$) de I dont le développement binaire commence par n chiffres donnés $a_1, \ldots, a_n$ est celui des nombres irrationnels y, tels que

$$\frac{a_1}{2} + \ldots + \frac{a_n}{2^n} < y < \frac{a_1}{2} + \ldots + \frac{a_n + 1}{2^n}.$$

Ces nombres sont les irrationnels d'un segment de longueur $\frac{1}{2^n}$; leur mesure est donc $\frac{1}{2^n}$.

L'ensemble des y, pour lesquels la fréquence des 1 dans les n premiers chiffres de y est

$$f_n = \frac{p}{n},$$

est formé de ceux contenant dans les n premiers chiffres, p chiffres 1 et $n-p$ chiffres zéro. La mesure de cet ensemble sera évidemment $C_n \frac{1}{2^n}$.

Soit ε un nombre positif arbitraire. La mesure $m(n, \varepsilon)$ de l'en-

(¹) Si l'on considère une loi de probabilité uniforme de x, on voit facilement que, la probabilité que le $n^{\text{ième}}$ chiffre de x soit égal à 1, pour n donné, est bien égale à $\frac{1}{2}$.

semble $M_n(\varepsilon)$ des irrationnels x pour lesquels $\left|f_n - \frac{1}{2}\right| \geqq \varepsilon$ est

$$m(n, \varepsilon) = \sum_p{}' C_n^p \frac{1}{2^n},$$

où Σ' est la sommation effectuée pour les valeurs de p telles que $\left|f_n - \frac{1}{2}\right| \geqq \varepsilon$

En posant $q = n - p$, cette inégalité peut s'écrire $\left|\frac{p-q}{2n}\right| \geqq \varepsilon$. Pour faire apparaître $p - q$, on va employer le procédé de Bertrand (p 85), et transformer l'identité

$$(\alpha + \beta)^n \equiv \sum_p C_n^p \alpha^p \beta^q$$

en repétant plusieurs fois sur elle l'opération $\alpha \frac{\partial}{\partial\alpha} - \beta \frac{\partial}{\partial\beta}$.

Écrivons au premier membre le résultat obtenu en posant $u = \alpha + \beta$, $v = \alpha - \beta$. On aura

$$nu^{n-1}v = \sum_p C_n^p (p - q)\, \alpha^p \beta^q,$$

$$nu^n + n(n-1)u^{n-2}v^2 = \sum_p C_n^p (p - q)^2 \alpha^p \beta^q,$$

$$(3n^2 + 2n)u^{n-1}v + n(n-1)(n-2)u^{n-3}v^3 = \sum_p C_n^p (p - q)^3 \alpha^p \beta^q,$$

$$\begin{aligned}(3n^2 - 2n)u^n + n(n-1)(6n-8)u^{n-2}v^2 \\ + n(n-1)(n-2)(n-3)u^{n-4}v^4 = \sum_p C_n^p (p-q)^4 \alpha^p \beta^q\end{aligned}$$

Faisons $\alpha = \beta = \frac{1}{2}$, d'où $v = 0$, $u = 1$, on aura

$$3n^2 - 2n = \sum_p C_n^p (p - q)^4 \frac{1}{2^n}.$$

Et, pour $\left|\frac{p-q}{2n}\right| \geqq \varepsilon$ ..

$$3n^2 > 3n^2 - 2n \geqq \left[\sum_p{}' C_n^p \frac{1}{2^n}\right] (2n\varepsilon)^4 = (2n\varepsilon)^4\, m(n, \varepsilon).$$

D'où

$$m(n, \varepsilon) < \frac{3}{16\varepsilon^4} \frac{1}{n^2}.$$

Estimons la mesure $m(\varepsilon)$ de l'ensemble $M(\varepsilon)$ des points y pour lesquels a lieu une infinité de fois l'inégalité $\left|\frac{p}{n} - \frac{1}{2}\right| \geqq \varepsilon$. Cet ensemble est, quel que soit n, compris dans l'ensemble $S_n(\varepsilon)$ des points appartenant à l'un au moins des ensembles $M_n(\varepsilon)$, $M_{n+1}(\varepsilon)$, . On a donc

$$\text{mes } M(\varepsilon) \leqq \text{mes } S_n(\varepsilon) \leqq m_n(\varepsilon) + m_{n+1}(\varepsilon) + \dots$$
$$\leqq \frac{3}{16\varepsilon^4}\left[\frac{1}{n^2} + \frac{1}{(n+1)^2} + \dots\right]$$

Ceci ayant lieu, quel que soit n, on a

$$\text{mes } M(\varepsilon) = 0$$

quel que soit $\varepsilon > 0$.

Or pour tout point y de l'ensemble M des points où $\frac{p}{n}$ ne converge pas vers $\frac{1}{2}$, il existe au moins une valeur de l'entier r telle que l'inégalité $\left|\frac{f}{n} - \frac{1}{2}\right| \geqq \frac{1}{r}$ ait lieu une infinité de fois Par suite, M appartient à la réunion de la suite dénombrable d'ensembles de mesures nulles $M(1)$, $M\left(\frac{1}{2}\right)$, $M\left(\frac{1}{r}\right)$, ... Donc M est de mesure nulle, ce qui démontre le théoreme de M. Borel.

M. Hausdorff observe que ce théorème est tres remarquable, qu'en particulier, il apparaît comme une extension plausible a l'infini de la loi des grands nombres.

Enfin, il ajoute qu'on peut préciser la rapidité de la convergence de f_n en démontrant que l'ensemble plus étroit des points ou $\left|f_n - \frac{1}{2}\right| n^{\theta}$ tend vers zéro avec n, garde aussi la mesure 1 pour toute valeur de $\theta < \frac{1}{2}$. Cette proposition a été précisée successivement de plus en plus par une suite d'auteurs arrivant à la forme dite du logarithme itéré, elle-même précisée à son tour de plus en plus Nous remettrons au volume « Compléments divers » l'étude de cette intéressante question.

Remarque. — Il est bien clair que si la variable aléatoire Y_n converge « au sens ordinaire » vers la variable aléatoire Y, elle converge aussi, *a fortiori*, « presque certainement » vers Y

Théorème. — *Si* X_n *converge « presque certainement » vers* X, X_n *converge aussi « en probabilité » vers* X

Soit e l'événement consistant en ce que X_n converge vers X, E_n l'événement consistant en ce que $|X_n - X| < \eta$, e_n l'événement consistant en ce qu'on a à la fois.

$$|X_n - X| < \eta, \qquad |X_{n+1} - X| < \eta,$$

Il est clair que, si e_n a lieu, e_{n+1} aussi. Donc

$$e_n \subset e_{n+1}$$

L'événement E consistant en ce qu'ait lieu l'un au moins des événements $e_1, e_2, \ldots, e_n, \ldots$, on a (p. 24)

$$\Pr \{E\} = \lim_{n \to \infty} \Pr \{e_n\}. \tag{167}$$

Or pour chaque épreuve où X_n converge vers X, l'un au moins des événements $e_1, e_2, \ldots$ a lieu. Autrement dit, quand e a lieu, E aussi. Donc

$$\Pr \{e\} \leqq \Pr \{E\}. \tag{167 bis}$$

Et, comme X_n converge « presque certainement » vers X

$$1 = \Pr \{e\} \leqq \Pr \{E\} \leqq 1$$

D'où

$$\lim_{n \to \infty} \Pr \{e_n\} = 1 \tag{168}$$

Or, puisque E_n a lieu quand e_n a lieu,

$$\Pr \{e_n\} \leqq \Pr \{E_n\} \tag{169}$$

On tire de (168) et (169)

$$\lim_{n \to \infty} \Pr \{E_n\} = 1$$

Ceci ayant lieu pour toute valeur de $\eta > 0$, la proposition est établie.

Il en résulte en particulier que :

Si une variable aléatoire Y_n *converge au sens ordinaire vers la variable aléatoire* Y, *elle converge aussi vers* Y « *en probabilité* », proposition obtenue précédemment par une autre voie.

La réciproque de cette dernière proposition n'est pas exacte, comme nous l'avons vu plus haut, page 203. N'est pas non plus exacte

la réciproque du théorème qui vient d'être démontré. C'est ce qui résulte de l'exemple de la page 203.

Si la convergence « presque certaine » est, d'après cela, plus stricte que la convergence « en probabilité », elle se trouve par contre équivalente à une autre nature de convergence dont nous allons parler : la convergence uniforme « en probabilité ».

Chemin faisant, nous rattacherons ces deux notions à une autre : la convergence forte, qui ne jouera qu'un rôle d'intermédiaire. M. Cantelli a montré explicitement, le premier, que le théorème de Bernoulli peut être précisé au moyen de cette notion. Soit f_n la fréquence d'un événement de probabilité constante p au cours de n épreuves, $Q_n(\eta)$ la probabilité pour que $|f_n - p| < \eta$, $\varpi_n(\eta)$ la probabilité pour que l'on ait *à la fois*, comme résultat d'une même épreuve,

$$|f_n - p| < \eta, \quad |f_{n+1} - p| < \eta, \quad \ldots$$

Il est clair que $\varpi_n(\eta) \leqq Q_n(\eta)$. Le théorème de Bernoulli affirme que $Q_n(\eta)$ tend vers l'unité quand n croît indéfiniment. M. Cantelli a démontré la proposition plus précise : $\varpi_n(\eta)$ tend aussi vers l'unité.

C'est ce mode de convergence que M. Khintchine appelle la convergence forte, au sens du Calcul des Probabilités, de f_n vers p.

Convergence forte. — Disons provisoirement d'une manière générale, qu'une variable aléatoire X_n converge fortement au sens du Calcul des Probabilités vers la variable aléatoire X, si, pour tout $\varepsilon > 0$, on a

$$\lim_{n \to \infty} \Pr [e_n^{(\varepsilon)}] = 1,$$

en appelant $e_n^{(\varepsilon)}$ l'événement consistant en ce qu'on a pour une même épreuve, à la fois, les inégalités

$$(170) \qquad |X_n - X| < \varepsilon, \quad |X_{n+1} - X| < \varepsilon, \quad \ldots$$

Il résulte de l'égalité (168) de la page précédente que, si une variable aléatoire X_n converge « presque certainement » vers une variable aléatoire X, X_n converge fortement au sens du Calcul des Probabilités vers X.

Seulement la convergence ainsi définie fait jouer un rôle de premier plan à des événements $e_n^{(\varepsilon)}$ *à double indice* et en particulier

laisse sans réponse la question de la comparaison des *limites* de X_n et des valeurs de X. Or on peut préciser et même de deux façons.

Convergence uniforme en probabilité. — Si l'on suppose que X_n converge fortement vers X, alors, étant donnés deux nombres $\varepsilon > 0$, $\eta > 0$, on peut trouver un certain nombre N tel qu'il y ait une probabilité $> 1 - \eta$ que l'on ait à la fois

$$|X_n - X|$$

pour toutes les valeurs de $n > N$.

Soit p un entier arbitraire. Appelons $e^{(p)}$ l'événement consistant dans le concours simultané des inégalités $|X_n - X| < \frac{1}{p}$ pour n supérieur ou égal à un certain entier N_p.

D'après ce qui précède, on peut choisir N_p de sorte que

$$\Pr |e^{(p)}| \quad 1 - \frac{1}{2^p} \qquad \text{et} \qquad N_p \quad N_{p-1}$$

L'événement E_r consistant dans le concours des événements $e^{(r+1)}$, $e^{(r+2)}$, ..., $e^{(r+p)}$, ... aura une probabilité $\geqq 1 - \frac{1}{2^r}$.

Il en sera, *a fortiori*, de même de tout événement consistant dans la réalisation d'une partie seulement des inégalités qui caractérisent E_r. On a donc, en particulier $\Pr. F_n \geqq 1 - \frac{1}{2^r}$ quand $N_{r+1} \leqq n < N_{r+2}$, en désignant par F_n la réalisation simultanée des inégalités

$$(171) \quad |X_n - X| < \varepsilon_n, \quad |X_{n+1} - X| < \varepsilon_{n+1}, \quad \ldots, \quad |X_{n+s} - X| < \varepsilon_{n+s}, \ldots$$

en posant $\varepsilon_n = \frac{1}{r+1}$, de sorte que

$$\varepsilon_{N_{r+1}} = \varepsilon_{1+N_{r+1}} = \varepsilon_{2+N_{r+1}} = \ldots = \varepsilon_{N_{r+2}-1} = \frac{1}{r+1},$$

$$\varepsilon_{N_{r+2}} = \varepsilon_{1+N_{r+2}} = \ldots = \frac{1}{r+2},$$

Quand n croît indéfiniment, il en est de même de r et par suite ε_n tend vers zéro et la probabilité de F_n tend vers l'unité.

Réciproquement, quand il en est ainsi, il est visible que X_n converge fortement vers X au sens provisoire. Nous arrivons ainsi à une seconde

définition de la convergence forte : X_n converge fortement vers X quand il existe *au moins une suite* de nombres ε_n *tendant vers zéro*, tels que la probabilité de la réalisation simultanée des inégalités (171) tend vers l'unité quand n croît.

On aurait pu augurer, *a priori*, que cette seconde définition [qui substitue aux ε égaux de (170) les ε_n tendant vers zéro de (171)] est plus stricte que la première. Nous venons de voir (au moyen d'un raisonnement qui nous a été communiqué par M. Cantelli) qu'en réalité ces deux définitions sont équivalentes.

D'ailleurs, en vertu de (171) et de l'hypothèse que $\lim\limits_{n \to \infty} \varepsilon_n = 0$, on voit que X_n converge uniformément vers X quand l'événement F_n a lieu et l'on a

$$\lim_{n \to \infty} \Pr \{ F_n \} = 1$$

Il en résulte en particulier que l'ensemble des épreuves où X_n converge au sens ordinaire vers X a une probabilité égale à l'unité, car cet ensemble comprend évidemment chaque F_n. Dès lors X_n converge « presque certainement » vers X.

D'ailleurs la réciproque est vraie, comme nous l'avons vu plus haut. Ainsi *la convergence forte au sens du Calcul des Probabilités est entièrement équivalente à la convergence qui a lieu « presque certainement »*. Mais nous avons obtenu quelque chose de plus. Nous avons montré que dans la convergence forte, le rôle tenu par les événements à double indice $e_n^{(\varepsilon)}$ peut être tenu par les événements à simple indice F_n. Et ceux-ci ont une signification beaucoup plus simple :

X_n converge uniformément pour chacun des événements F_n dont, en outre, la probabilité tend vers 1 quand n croît. Il est dès lors naturel de dire que dans ces conditions X_n *converge uniformément « en probabilité »* vers X. Il est d'ailleurs clair que, réciproquement, la convergence uniforme « en probabilité » entraîne, par définition même, la convergence forte au sens du Calcul des Probabilités.

Finalement, nous avons établi que : la condition nécessaire et suffisante pour que la variable X_n converge uniformément « en probabilité » vers X est que X_n converge « presque certainement » vers X.

Nous voyons que *si l'on porte son attention sur la propriété de*

convergence ordinaire, il y a lieu de donner à la convergence forte au sens du Calcul des Probabilités la forme équivalente de la convergence qui a lieu « presque certainement ». *Si*, au contraire, *c'est de l'uniformité de la convergence qu'on a besoin, il sera préférable de donner à la convergence forte au sens du Calcul des Probabilités la forme équivalente de la convergence uniforme « en probabilité »*.

Remarques. — 1° La seconde définition de la convergence forte peut s'exprimer sous une forme plus frappante. Appelons F l'événement consistant dans la réalisation simultanée à partir d'un certain rang des inégalités

$$(171\ bis) \qquad |X_1 - X| < \varepsilon_1, \quad |X_2 - X| < \varepsilon_2, \quad \ldots, \quad |X_n - X| < \varepsilon_n, \quad \ldots.$$

Il est clair que F consiste dans la réalisation de F_1, ou F_2, ou ..., ou F_n, ou ...; et comme F_n implique F_{n+1}, on a (p. 24),

$$\Pr F = \lim_{n \to \infty} \Pr F_n$$

Dès lors, on a

$$\Pr F = 1 \qquad \text{ou} \qquad < 1,$$

suivant que l'on a

$$\Pr \lim_{n \to \infty} F_n = 1 \qquad \text{ou} \qquad = 1$$

Donc, dire que X_n *converge uniformément en probabilité vers* X, c'est dire aussi qu'*il existe au moins* UNE *suite convergeant vers zéro de nombres certains* $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots$, *telle que soit égale à l'unité la probabilité de la réalisation simultanée des inégalités* (171 *bis*) *à partir d'un certain rang* (*variable éventuellement d'une épreuve à l'autre*).

Quand on veut démontrer une propriété de la convergence en question, cette définition est plus commode que la première définition proposée. Quand on veut démontrer qu'une suite converge fortement, il sera plus facile de prouver la condition qui caractérise cette première définition. Enfin, rappelons que ces deux définitions sont elles-mêmes équivalentes à celles de la convergence presque certaine et de la convergence uniforme en probabilité.

2° On pourrait songer à exprimer ce qui précède en disant que si X_n converge « presque certainement » vers X, X_n converge même « presque sûrement uniformément ».

Toutefois, il y aurait inconvénient à le faire pour la raison suivante. Quand nous disons que X_n converge « presque certainement » vers X, nous entendons par là qu'il y a un événement déterminé e de probabilité égale à l'unité, tel que la convergence ordinaire ait lieu quand e se produit.

Par analogie, il est naturel de dire qu'une variable aléatoire Y_n converge « presque toujours uniformément » vers la variable aléatoire Y lorsqu'il existe un événement déterminé e' de probabilité égale à l'unité tel que X_n converge uniformément vers X dans l'ensemble des épreuves où e' a lieu. Or cette condition est plus stricte que celle qui a été établie dans le dernier théorème. Et l'on peut donner *un exemple d'une variable aléatoire Z_n qui converge « presque certainement » et même toujours vers la variable aléatoire Z et qui ne converge pas uniformément « presque sûrement »*. Autrement dit, pour cette suite particulière de variables, il n'y a aucun événement e' de probabilité égale à l'unité tel que X_n converge vers X uniformément quand e' a lieu.

Reprenons en effet l'exemple de la page 203, mais en choisissant autrement la fonction $f_n(x)$. Prenons encore pour $y = f_n(x)$ une ligne polygonale, mais dont les sommets ont les ordonnées

$$f_n(0) = 0, \qquad f_n\left(\frac{1}{n}\right) = 1, \qquad f_n\left(\frac{2}{n}\right) = 0, \qquad f_n(1) = 0$$

Alors X_n converge toujours vers zéro, Soit maintenant e' un événement de probabilité égale à l'unité. Si X_n convergeait uniformément quand e' a lieu, alors, pour tout p entier, il y aurait un nombre N tel que $|X_n - X| = |X_n| < \frac{1}{p}$ partout sur e' pour $n > N$. Or $f_{N+1}(x) \geqq \frac{1}{2}$, par exemple, sur un intervalle de longueur $\frac{1}{N+1}$. Si l'on suppose $p > 2$, le hasard ne pourrait déterminer un point x appartenant à la fois à cet intervalle et à e' pour lequel $f_{N+1}(x) < \frac{1}{p}$. Donc la probabilité de e' serait inférieure à $1 - \frac{1}{N+1}$, alors qu'on suppose cette probabilité égale à l'unité.

Il y a donc lieu de distinguer la convergence uniforme « presque sûre » de la propriété qui vient d'être établie précédemment. Mais il reste légitime de continuer à décrire celle-ci comme une convergence uniforme « en probabilité ».

Ainsi nous dirons qu'une suite de variables aléatoires X_n *converge uniformément « en probabilité »* vers la variable aléatoire X, si à tout nombre $\eta > 0$ correspond un événement F_η de probabilité $> 1 - \eta$ et tel que X_n converge uniformément vers X quand a lieu l'événement E_η.

Nous avons montré que *la convergence uniforme « en probabilité » est équivalente à la convergence « presque certaine »*.

Suite extraite d'une suite donnée. — La seconde et par suite la première sorte de convergence sont liées à la convergence en probabilité comme le montre le raisonnement fait page 168, pour établir que le critère de M. Slutsky est suffisant. Avec les notations du même raisonnement, on voit, en effet, qu'on a établi, chemin faisant, la possibilité d'extraire d'une suite $X_1, X_2, \ldots, X_n, \ldots$ qui converge en probabilité, une suite $X_{\nu_1}, X_{\nu_2}, \ldots$ qui converge (vers une variable X bien définie) quand a lieu un événement H avec Pr. H = 1.

Nous allons déduire de l'observation précédente la conséquence suivante :

Lorsqu'une suite de variables aléatoires X_n, tend « en probabilité » vers une variable aléatoire X, on peut extraire de cette suite, une suite de variables Y_r, satisfaisant aux deux conditions suivantes : 1° il y a une probabilité nulle que cette suite des Y_r diverge, 2° quel que soit ε positif, il y a un événement E_ε dont la probabilité est $> 1 - \varepsilon$ et qui est tel que, dans l'ensemble des épreuves où cet événement a lieu, la suite des Y_r converge uniformément. [On pourrait traduire brièvement cette seconde condition en disant qu'il y a une probabilité nulle que la série des Y_r ne converge pas uniformément, mais ce serait au risque de malentendus possibles que nous avons précisés page 232. Il vaut mieux exprimer (comme à la même page) la propriété 2° en disant : la suite des Y_r converge uniformément « en probabilité ».]

En effet, la suite des X_n devra d'abord satisfaire au critère de convergence de Cauchy. On pourra donc en extraire comme plus haut la suite des Y_r. Cette suite des Y_r converge « en probabilité » vers une certaine variable Y ; et Y et X sont « presque toujours » égales, puisque ce sont à la fois les limites « en probabilité » des Y_r. Soit ε_{r-2} le premier des nombres décroissants $\varepsilon_1, \varepsilon_2, \ldots$ $\left(\text{avec } \varepsilon_r = \frac{1}{2^r}\right)$

qui est $< \varepsilon$. Appelons E_ε l'évenement consistant dans la réalisation simultanée des inégalités.

$$|Y_r - X| < \varepsilon_{r-1}, \quad |Y_{r+1} - X| < \varepsilon_r, \quad |Y_{r+2} - X| < \varepsilon_{r+1}, \ldots$$

On a vu que les probabilités de ces inégalités sont respectivement supérieures à $1 - \varepsilon_{r-1}$, $1 - \varepsilon_r$, ... Alors la probabilite de E_ε est $> 1 - \varepsilon_{r-1} - \varepsilon_r - \varepsilon_{r+1} - \ldots = 1 - \varepsilon_{r-2} > 1 - \varepsilon$.

La suite des Y_r converge uniformément vers X dans l'ensemble des épreuves où se produit l'événement E_ε. C'est-à-dire que si η est un nombre positif quelconque, il y a un entier M tel que $|Y_s - X| < \eta$, pour $s > M$, M étant le même pour toutes les épreuves ou E_ε a lieu. Il suffit en effet de prendre $\varepsilon_M < \eta$.

La condition 2° est ainsi établie. Remarquons en outre que dans toutes les épreuves où l'événement E_ε a lieu, la suite des Y_r converge. La probabilité que la suite des Y_r converge étant au moins égale à la probabilité de l'événement E_ε, sera $> 1 - \varepsilon$. La probabilité de la convergence des Y_r étant celle d'un événement indépendant de ε sera donc égale à l'unité, ce qui établit la condition 1°.

Nous avons déjà fait observer qu'on ne peut employer le même raisonnement pour déduire de la condition 2°, énoncée sous la forme où nous l'avons démontrée, que la probabilité de l'uniformité de la convergence est égale à l'unité. Pour une épreuve déterminée, la convergence ou la divergence de la suite des X_n déterminés par cette épreuve est nettement caractérisée. Il n'en est pas de même de l'uniformité de la convergence, propriété collective et non individuelle.

Dans certains cas, la suite des Y_r se confondra avec la suite des X_n. On pourrait se demander *s'il n'en est pas toujours ainsi*. L'exemple de la page 203 montre qu'*il n'en est rien*. Cet exemple montre même qu'une suite de variables aléatoires X_n peut converger « en probabilité » vers une variable X sans jamais converger vers X au sens ordinaire, même pour une seule épreuve. [De même qu'une suite de fonctions mesurables peut converger *en mesure* vers $f(x)$ et pourtant ne converger vers $f(x)$ pour aucune valeur de x.]

La convergence presque certaine n'est pas compatible avec l'existence d'une distance. — Nous avions démontré dans notre mémoire de Calcutta, que la convergence ordinaire et la convergence presque partout d'une suite de fonctions ne peuvent s'exprimer d'aucune

manière par l'intermédiaire d'une distance. On peut en déduire des conséquences analogues en ce qui concerne la convergence ordinaire et la convergence presque certaine des variables aléatoires. On pourra même en faire ensuite la démonstration directe.

Démonstration indirecte. — Il s'agit donc de démontrer d'abord indirectement que si l'on considère la famille $\mathcal{F}$ des variables aléatoires définies sur une même catégorie C d'épreuves, *il est impossible d'attribuer quelle que soit* C *à tout couple de variables aléatoires* X, Y *de la famille* $\mathcal{F}$ *une distance* (X, Y) *vérifiant les première et troisième conditions* de la page 193, *et telle que* la condition nécessaire et suffisante, pour que (X_n et X appartenant à $\mathcal{F}$) X_n converge *presque certainement* vers X, est que la distance (X_n, X) tende vers zéro.

Supposons que la catégorie C considérée soit la catégorie C_0 qui consiste à déterminer au hasard la position d'un point (d'abscisse Z) sur le segment (0, 1) et que le hasard répartisse uniformément les choses sur ce segment. De sorte que la probabilité que $\alpha \leq Z \leq \beta$ soit égale à $\beta - \alpha$.

Toute variable aléatoire Y définie sur cette catégorie d'épreuves sera telle que la valeur de Y soit déterminée par le résultat de l'épreuve, résultat qui est entièrement caractérisé par la valeur de Z. Donc Y sera une fonction certaine $f(Z)$, de la variable aléatoire Z. Admettons que le théorème que nous avons en vue soit inexact. Dès lors il existera une distance pour deux variables aléatoires quelconques $f(Z)$, $g(Z)$ appartenant à la famille $\mathcal{F}$ considérée. Ce sera un nombre $(f, g) \geqq 0$ déterminé par les fonctions f et g. Il ne devra être nul que si $f(Z) - g(Z)$ est presque certainement nul, c'est-à-dire que si la probabilité que $f(Z) - g(Z)$ soit $\neq 0$ est nulle, c'est-à-dire si la mesure de l'ensemble des Z, où $f(Z) - g(Z)$ est $\neq 0$, est nulle, c'est-à-dire enfin si $f(Z)$ et $g(Z)$ sont égales presque partout. De même, on aura évidemment

$$(f, g) \leqq (f, h) + (h, g)$$

Et enfin dire que (f, f_n) tend vers zéro, c'est dire que $f_n(Z) - f(Z)$ converge presque certainement vers zéro, c'est-à-dire que $f_n(x) - f(x)$ converge presque partout vers zéro. On pourrait donc définir une distance compatible avec la convergence presque partout des fonctions d'une variable, ce qui est impossible [Fréchet, note (1), p. 161].

On démontrerait de même qu'il est impossible de définir pour toute catégorie C d'épreuves une distance de deux variables aléatoires quelconques définies sur C, de façon que cette distance puisse suffire à exprimer la *convergence ordinaire* (p. 202) de toute suite de ces variables aléatoires.

Démonstration directe. — La démonstration précédente porte sur une catégorie particulière C_0 d'épreuves et renvoie à une démonstration concernant les fonctions certaines.

On peut donner une démonstration directe portant sur toute catégorie C d'épreuves tout au moins quand C, comme C_0, comporte un ensemble non dénombrable d'épreuves possibles distinctes. Considérons la famille $\mathcal{F}$ des variables aléatoires définissables sur C et supposons qu'il existe pour C une distance compatible avec, par exemple, la convergence ordinaire. Soit X_0 une variable déterminée et $\mathcal{F}_\varepsilon$ l'ensemble des variables aléatoires X de $\mathcal{F}$ telles que la distance (X, X_0) soit $< \varepsilon$. Appelons T_ε une variable aléatoire définie sur C par la condition que, pour chaque épreuve de C, T_ε soit la borne supérieure des valeurs, pour cette épreuve, de $|X_0 - X|$ pour tous les X de $\mathcal{F}_\varepsilon$. Il est clair que $\mathcal{F}_\varepsilon$ appartient à $\mathcal{F}_{\varepsilon'}$ si $\varepsilon < \varepsilon'$ et que, par suite, $T_\varepsilon \leqq T_{\varepsilon'}$. La fonction monotone non décroissante T_ε étant $\geqq 0$, tend vers une limite T_0 quand ε tend vers zéro; (il n'est pas impossible que T_ε soit infini.) Je dis que T_0 est nul, c'est-à-dire que $\lim\limits_{\varepsilon \to 0} T_\varepsilon = 0$, d'où résultera en particulier que T_ε est fini pour ε assez petit $(\varepsilon < \varepsilon_0)$. En effet, si $T_0 \neq 0$ pour une épreuve déterminée E, et si l'on prend alors $\varepsilon_n = \frac{1}{n}$, et η tel que $T_0 > \eta > 0$, il y aurait une variable aléatoire X_n définie sur C, telle que $(X_0, X_n) < \varepsilon_n$ et que $|X_n - X_0| > \eta$ pour l'épreuve E. Alors puisque (X_n, X_0) tend vers zéro avec ε_n, X_n doit par hypothèse, converger certainement vers X_0 quelle que soit l'épreuve considérée. Or, pour l'épreuve E, on a $|X_n - X_0| > \eta > 0$, ce qui est contradictoire. Ainsi T_0 doit être nul.

Considérons maintenant une suite infinie $E_1, E_2, \ldots,$ d'épreuves distinctes appartenant à C et une suite infinie de variables aléatoires X_n de $\mathcal{F}$ telles que X_n soit égal à X_0 sauf pour l'épreuve E_n. Il est clair que X_p converge vers X_0 quand p croît, pour toute épreuve de C. On doit donc avoir

$$\lim_{p \to \infty} (X_p, X_0) = 0.$$

Or puisqu'il existe un nombre $\omega > 0$ tel que la valeur de T_ω soit finie pour l'épreuve E_n, on pourra prendre, pour valeur de X_n pour l'épreuve E_n, une valeur finie telle que $|X_n - X_0| > T_\omega$ pour cette épreuve E_n. C'est possible, car T_ω est déterminé d'avance, indépendamment du choix des X_n parmi les variables aléatoires définies sur C. Dans ces conditions, on aura nécessairement $(X_n, X_0) \geq \omega$ ce qui aboutirait à une contradiction si ω est indépendant de n. Il reste à établir l'existence d'un nombre ω indépendant de n. Si un tel nombre n'existait pas, alors pour toute valeur de $\varepsilon > 0$, il y aurait une valeur N de n telle que T_ε ne soit pas fini pour l'épreuve E_N. Et pour que la démonstration précédente ne fut pas applicable, il faudrait que ceci eut lieu quelle que soit la suite d'épreuves distinctes $E_1, E_2, \ldots, E_n, \ldots$ appartenant à C. Or, on sait que, pour toute épreuve E de C, T_ε est fini pour ε assez petit. Soit $\varepsilon_r = \frac{1}{r}$. Les épreuves pour lesquelles T_{ε_r} reste fini ne peuvent être en nombre fini quel que soit l'entier r, sans quoi elles formeraient quand r varie un ensemble dénombrable D. Or on a vu que, pour toute épreuve de C, T_{ε_r} est fini pour r assez grand donc E appartient à D. Dès lors la catégorie C serait dénombrable. Si, comme cela a lieu par exemple dans la catégorie C_0 considérée plus haut, la catégorie C est non dénombrable, il faut donc supposer qu'il y a une valeur r_0 de r telle que les épreuves pour lesquelles T_{ε_r} est fini, soient en nombre infini. On pourra donc prendre $\omega = \frac{1}{r_0}$, et prendre pour $E_1, \ldots, E_n, \ldots$ une infinité des épreuves pour lesquelles T_ω est fini.

Cas exceptionnel. — Non seulement la démonstration, mais le résultat obtenu ci-dessus cessent d'être valables quand la catégorie C considérée ne comporte qu'un ensemble dénombrable d'épreuves distinctes $E_1, E_2, \ldots$. Dans ce cas chaque variable aléatoire X définie sur C ne peut prendre qu'un ensemble dénombrable de valeurs $x_1, x_2, \ldots$. Alors, *on peut* définir une distance compatible avec la convergence ordinaire; il suffit d'appeler distance de deux variables aléatoires X, Y (prenant les valeurs x_k, y_k pour l'épreuve E_k), la somme de la série convergente

$$(X, Y) = \sum_{k=1}^{k=\infty} \frac{1}{k!} \frac{|x_k - y_k|}{1 + |x_k - y_k|}.$$

On démontre comme pour l'espace E_ω (E. A., p. 81) que cette quantité représente pour les variables aléatoires définies sur C, une distance compatible avec la convergence certaine.

Conséquence. — Nous avons obtenu trois échelles de convergence de plus en plus strictes :

la convergence « en probabilité » ;

la convergence « presque certaine » ou la convergence uniforme « en probabilité »,

la convergence uniforme « presque certaine ».

Premier critère de convergence presque certaine. — Nous venons de montrer (p. 233) que d'une suite de variables aléatoires qui converge « en probabilité », on peut tirer une suite de variables qui converge « presque certainement » ou, ce qui revient au même, qui converge uniformément « en probabilité ».

Il est interessant de déterminer des cas très généraux où l'on soit assuré qu'il en est ainsi pour la suite primitive elle-même. Tel est le cas suivant qui comprend, comme nous le verrons, le cas de Bernoulli :

Pour qu'une suite de variables aléatoires X_n converge « presque certainement » vers une variable aléatoire X, il suffit que, pour au moins une valeur de l'ordre positif s, la série

$$M_1^{(s)} + \ldots + M_n^{(s)} + \ldots$$

soit convergente, en appelant $M_n^{(s)}$ la valeur moyenne de $|X_n - X|^s$.

En effet, d'après l'inégalité de Bienaymé, on a

$$\Pr\left[|X_n - X| < \varepsilon\right] \geqq 1 - \frac{M_n^{(s)}}{\varepsilon^s}.$$

On a donc, avec les notations de la page 228,

$$1 \geqq \Pr\, e_n^{(\varepsilon)} \geqq 1 - \frac{M_n^{(s)} + M_{n+1}^{(s)} + \ldots}{\varepsilon^s};$$

d'où

$$\lim_{n=\infty} \Pr.\, e_n^{(\varepsilon)} = 1$$

pour tout ε positif. Ainsi X_n converge fortement et par suite « presque certainement » vers X.

Théorème de Cantelli. — En particulier, prenons $X = 0$ et $X_n = v_n - \overline{v_n}$ en posant

$$v_n = \frac{Y_1 + \dots + Y_n}{n}$$

et en appelant $Y_1, Y_2, \dots, Y_n, \dots$ une suite quelconque de variables aléatoires *indépendantes*. *Pour que $v_n - \overline{v_n}$ converge « presque certainement » vers zéro, il suffit que les écarts moyens ν_k d'ordre 4 des Y_k soient non seulement finis mais bornés dans leur ensemble.* Car, *s'il existe un nombre* B *supérieur* à tous les ν_k on aura

$$(172)\qquad M_n^{(4)} = \frac{1}{n^4}\left[\sum_k \overline{(Y_k - \overline{Y_k})^4} + 6\sum_{k,h} \overline{(Y_k - \overline{Y_k})^2\,(Y_h - \overline{Y_h})^2}\right]$$
$$\leq \frac{1}{n^4}\left[n B^4 + 6\,\frac{n(n-1)}{2}\,B^4\right]$$

Par suite, la série $\Sigma M_n^{(4)}$ est bien convergente. *C'est ce qui aura lieu, en particulier, si les Y_k sont bornés dans leur ensemble.* Car alors il en sera de même des $\overline{Y_k}$ et par suite des ν_k.

Dans le cas où la suite des valeurs moyennes $\overline{Y_n}$ des variables indépendantes Y_n converge vers une limite finie V, il suffit que les écarts moyens d'ordre 4 des Y_k soient bornés dans leur ensemble pour que la moyenne arithmétique $v_n = \frac{Y_1 + \dots + Y_n}{n}$ converge uniformément « en probabilité » vers V. C'est, sous une forme un peu simplifiée, le théorème de M. Cantelli (6, p. 43-44). Nous verrons plus loin, pages 242-243, qu'il reste valable sous des conditions plus générales encore.

Lorsque les Y_n indépendants, sont bornés dans leur ensemble et quand, en outre, $\overline{Y_n}$ tend vers V, alors cela suffit pour que v_n converge encore uniformément « en probabilité » vers V.

Ce cas particulier comprend celui de Bernoulli, où Y_k est égal à 0 ou 1 et où v_n représentant la fréquence f_n dans n épreuves d'un événement de probabilité constante p, on a $\overline{v_n} = p$. Dès lors est établi, comme conséquence particulière d'un raisonnement dû à M. Cantelli (5), que *si f_n est la fréquence au cours de n épreuves d'un événement fortuit de probabilité constante p, f_n converge uniformément vers p au sens du Calcul des Probabilités, quand n croît.* Nous

savons d'ailleurs que cet énoncé, distinct du corollaire donné page 217, du théorème de M. Borel, lui est équivalent en vertu de l'équivalence établie page 230 de la convergence uniforme (au sens du Calcul des Probabilités) de M. Cantelli et de la convergence presque certaine de M. Borel. Le raisonnement de M. Cantelli est plus simple que celui de M. Borel, il a fourni le résultat plus général ci-dessus et va s'appliquer sans effort dans un instant au cas de Poisson. Par contre, celui de M. Borel a l'avantage de se prêter mieux peut-être à une étude plus précise du comportement de la fréquence telle qu'elle a été entreprise par MM. Khintchine, Kolmogoroff et Paul Lévy. Cependant, comme nous le verrons quelques lignes plus loin, M. Cantelli avait déduit lui-même de son premier raisonnement certaines précisions sur ce comportement et signalé qu'on pourrait améliorer ces précisions.

Dans le cas, dit de Poisson, où la probabilité de l'événement considéré à la $n^{\text{ième}}$ épreuve est un nombre p_n dépendant de n, la valeur moyenne de ν_n est $\overline{\nu_n} = \frac{p_1 + \dots + p_n}{n} = p^{(n)}$. Mais puisqu'ici les X_k étant encore 0 ou 1 sont bornés dans leur ensemble, la *déviation $\nu_n - p^{(n)}$ de la fréquence ν_n à partir de sa moyenne $p^{(n)}$ converge encore presque certainement vers zéro.* C'est l'un des exemples qui ont conduit M. Slutsky à introduire une notion nouvelle, celle de l'équivalence asymptotique, sur laquelle nous reviendrons.

Dans le cas considéré par M. Cantelli, celui où $p^{(n)}$ a une limite p, la fréquence ν_n converge fortement vers p.

Forme initiale du théorème de Cantelli. — Afin d'arriver plus vite au résultat principal, nous avons un peu écourté plus haut (page précédente) le raisonnement de M. Cantelli. Celui-ci fournissait une précision supplémentaire intéressante. Reprenons les notations de la page précédente en remplaçant ε par ε_n, on a

$$\Pr\{|X_n - X| < \varepsilon_n\} \geqq 1 - \frac{M_n^{(4)}}{\varepsilon_n^4},$$

et en désignant par K_n l'événement consistant dans la réalisation simultanée des inégalités

(170 *bis*) $$|X_n - X| < \varepsilon_n, \quad |X_{n+1} - X| < \varepsilon_{n+1}, \quad \dots,$$

on a

$$\Pr K_n \geqq 1 - \left(\frac{M_n^{(4)}}{\varepsilon_n^4} + \frac{M_{n+1}^{(4)}}{\varepsilon_{n+1}^4} + \dots \right)$$

Il est clair que si l'on choisit les ε_n, de sorte que $\sum_n \frac{M_n^4}{\varepsilon_n^4}$ soit une série convergente, on aura

$$\lim_{n \to \infty} \Pr K_n = 1 \tag{173}$$

Or, en vertu de (172), $\frac{M_n^4}{\varepsilon_n^4} \leqq \frac{3B^4}{n^2\varepsilon_n^4}$. Pour assurer l'égalité (173), il suffit de prendre les ε_n, de sorte que $\frac{1}{n^2\varepsilon_n^4}$ soit un infiniment petit de l'ordre de $\frac{1}{n^{1+\theta}}$ où $0 < \theta < 1$, c'est-à-dire de sorte que ε_n soit un infiniment petit de l'ordre de $\frac{1}{n^{\frac{1-\theta}{4}}}$ On en déduit deux résultats distincts, dont l'ensemble constitue un énoncé plus proche que le précédent du théorème réellement démontré par M. Cantelli :

a. Si des variables indépendantes Y_n sont telles que la suite de leurs valeurs moyennes $\overline{Y}_n$ ait une limite V, il suffit que leurs écarts moyens d'ordre 4 soient bornés pour qu'il existe une suite de nombres certains $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots$ tendant vers zéro telle que la probabilité de la réalisation simultanée K_n des inégalités (170 *ter*)

$$\left|\frac{Y_1 + \ldots + Y_n}{n} - V\right| < \varepsilon_n, \quad \left|\frac{Y_1 + \ldots + Y_{n+1}}{n+1} - V\right| < \varepsilon_{n+1} \quad \ldots$$

converge vers l'unité quand n croît

b. Dans les hypothèses précédentes, on peut prendre pour ε_n tout infiniment petit d'ordre numérique $\left(\text{en } \frac{1}{n}\right)$ inférieur à $\frac{1}{4}$.

En terminant sa démonstration, M. Cantelli faisait observer qu'on pourrait, en suivant une méthode analogue, préciser le second résultat. Et, en effet, nous avons vu, tout au moins dans le cas de Bernoulli, que la méthode de M. Hausdorff permettrait de prendre pour ε_n un infiniment petit d'ordre inférieur à $\frac{1}{n^{\frac{1}{2}}}$. Et nous avons indiqué que d'autres méthodes plus récentes permettent d'aller plus loin.

Mais revenons au résultat *a*. En le généralisant, il conduit à la seconde définition de la convergence forte qui a été formulée page 230.

Conditions suffisantes plus larges pour la convergence « presque certaine ». — Nous avons tenu à donner la condition suffisante de convergence « *presque certaine* » énoncée plus haut (p. 239) parce

que c'est celle qui a suffi à M. Cantelli pour établir l'existence si importante de la convergence forte dans le cas de Bernoulli

Mais on peut donner des critères plus larges. Considérons d'abord une suite $X_1, X_2, \ldots, X_n, \ldots$ dont nous nous demandons si elle converge ou non vers la variable aléatoire X. L'événement E consistant dans la non-convergence de X_n vers X consiste aussi dans la réalisation de l'un au moins des événements E_r, E_r consistant lui-même dans la réalisation d'une infinité des événements $E_{r,n}$ définis par

$$|X_n - X| \geqq \frac{1}{r}.$$

On a donc (p. 24)

$$\text{Pr. } E \leqq \sum_r \text{Pr. } E_r,$$

et, d'autre part, si la série

$$\sum_n \text{Pr} \left\{ |X_n - X| \geqq \frac{1}{r} \right\} \tag{173}$$

est convergente, on a vu (p. 26) que la probabilité de E_r est nulle Donc Pr. $E = 0$ *si cette série* (173) *est convergente pour tout* $r > 0$, X_n *converge presque certainement vers* X. Nous savons d'ailleurs qu'il suffit pour la convergence « en probabilité » de X_n vers X que le terme général de la série (173) tende vers zéro avec $\frac{1}{n}$ pour tout $r > 0$.

En estimant Pr. $\left\{ |X_n - X| > \frac{1}{r} \right\}$ au moyen de l'inégalité de Bienaymé d'ordre s, on retrouve le critère de convergence moins étendu énoncé page 238.

M Borel semble avoir été le premier à donner un exemple d'une classe d'événements aléatoires telle que la probabilité de chacun d'eux ne puisse être que 0 ou 1 et que la distinction soit liée à la convergence ou à la divergence d'une série de nombres certains associés à ces événements. Ce résultat établi, page 27, peut être appliqué à l'étude de la convergence d'une suite de variables aléatoires. Pour pouvoir l'utiliser, supposons maintenant que les différences $X_n - X$ soient des variables aléatoires *indépendantes*. Alors les événements $E_{r,1}$, $E_{r,2}, \ldots, E_{r,n}, \ldots$ sont indépendants et la divergence de la série (173) ne peut se produire comme on l'a vu, page 27 que si $P(E_r) = 1$.

D'ailleurs

$$\Pr\left\{|X - X_n| \geqq \frac{1}{r}\right\} \leqq \Pr\left\{|X - X_n| \geqq \frac{1}{r+1}\right\}.$$

Donc si la série (173) n'est pas convergente pour tout entier r, elle est divergente à partir d'une certaine valeur R de r et, par suite, $\Pr.\, E_r = 1$ pour $r \geqq R$. Or

$$1 \geqq \Pr\, E \geqq \Pr\, E_r = 1, \qquad \text{d'où} \qquad \Pr\, E = 1$$

Dès lors : *si les différences* $|X_n - X|$ *sont indépendantes, la probabilité de la convergence de* X_n *vers* X *ne peut etre égale qu'à* 0 *ou* 1. *Pour que* X_n *converge presque certainement vers* X, *il faut et il suffit que la série*

$$\sum_n \Pr\left\{|X_n - X| > \frac{1}{r}\right\}$$

converge pour toute valeur entière de r.

Revenons maintenant au cas où les variables aléatoires de la suite $X_1, X_2, \ldots$ dont on veut étudier la convergence sont quelconques (indépendantes ou non).

Si la limite X n'est pas connue d'avance, on aura à faire intervenir les quantités

$$\Pr\left\{|X_n - X_{n+p}| > \varepsilon\right\}$$

Appelons $\varpi_n(\varepsilon)$ la borne supérieure de ces quantités pour $p = 1, 2, \ldots$.

Pour que X_n converge presque certainement, il faut d'abord que X_n converge en probabilité et, d'après le critère de Slutsky, il faut que $\varpi_n(\varepsilon)$ tende vers zéro avec $\frac{1}{n}$ pour chaque ε fixe.

Inversement, supposons que $\varpi_n(\varepsilon)$ converge vers zéro pour chaque ε positif. Alors X_n converge au moins en probabilité vers une limite X. En vertu du lemme de la page 162, on a

$$(174)\quad \Pr\left\{|X - X_n| > 2\varepsilon\right\} \leqq \Pr\left\{|X - X_{n+p}| > \varepsilon\right\} + \Pr\left\{|X_n - X_{n+p}| > \varepsilon\right\}.$$

En faisant croître p vers l'infini, on a donc

$$\Pr.\left\{|X - X_n| > 2\varepsilon\right\} \leqq \varpi_n(\varepsilon)$$

Dès lors, *si la série*

$$(175)\qquad \sum_n \varpi_n(\varepsilon)$$

est convergente pour tout $\varepsilon > 0$, la suite des X_n converge « presque certainement »

On peut même préciser encore ce résultat

Considérons la suite

$$\Pr\{|X_n - X_{n+1}| > \varepsilon\}, \quad \Pr\{|X_n - X_{n+2}| > \varepsilon\}, \ldots$$

et soit $\varpi'_n(\varepsilon)$ la plus petite de ses limites, c'est un nombre $\leq \varpi_n(\varepsilon)$. Si dans la formule (174) on fait croître p, non par valeurs quelconques mais par valeurs convenables, on aura

$$\Pr\{|X - X_n| > \varepsilon\} \leq \varpi'_n(\varepsilon).$$

Dès lors, si, quel que soit ε positif, $\varpi_n(\varepsilon)$ tend vers zéro et la série

$$\sum_n \varpi'_n(\varepsilon) \tag{176}$$

converge, il y aura encore convergence presque certaine de X_n vers X.

Formule de Kolmogoroff. — Une formule utile a été donnée par M. Kolmogoroff pour exprimer la probabilité P de la convergence d'une suite $X_1, X_2, \ldots$ de variables aléatoires *quelconques*. L'événement E consistant dans cette convergence est identique d'après le critère de Cauchy au concours des événements $E(\varepsilon)$ consistant chacun dans l'existence d'un entier n tel que

$$|X_p - X_m| < \varepsilon \qquad \text{pour } p \text{ et } m \text{ supérieurs ou égaux à } n$$

Comme

$$E(\varepsilon) \subset E(\varepsilon') \qquad \text{pour } \varepsilon < \varepsilon'$$

on aura (p. 24)

$$P = \lim_{\varepsilon \to 0} \Pr E(\varepsilon).$$

Mais $E(\varepsilon)$ consiste dans la réalisation de l'un au moins des événements $E_n(\varepsilon)$ consistant en ce que l'on ait $|X_p - X_m| < \varepsilon$ pour p et m supérieurs ou égaux à n. Et comme

$$E_{n+1}(\varepsilon) \supset E_n(\varepsilon),$$

on a donc (p. 24)

$$\Pr. E(\varepsilon) = \lim_{n \to \infty} \Pr E_n(\varepsilon).$$

Enfin $E_n(\varepsilon)$ consiste dans la réalisation, pour $m = n, n+1, \ldots$

simultanément, des événements $E_{n,m}(\varepsilon)$ consistant chacun dans la réalisation simultanée des inégalités $|X_p - X_r| < \varepsilon$, où r et p prennent indépendamment toutes les valeurs $n, n+1, \ldots, m$. Et comme

$$E_{n,m+1}(\varepsilon) \subseteq E_{n,m}$$

on a (p. 24)

$$\Pr E_n(\varepsilon) = \lim_{m \to \infty} \Pr E_{n,m}(\varepsilon)$$

D'où

$$P = \lim_{\varepsilon \to 0} \left\{ \lim_{n \to \infty} \left[\lim_{m \to \infty} \Pr E_{n,m}(\varepsilon) \right] \right\} \tag{177}$$

On peut obtenir un résultat parfois plus commode en appelant $F_{n,m}(\varepsilon)$ la probabilité de la réalisation simultanée des inégalités

$$|X_n - X_r| < \varepsilon \quad \text{pour } r = n+1,\ n+2,\ \ldots,\ m$$

On a évidemment

$$E_{n,m}(\varepsilon) \subseteq F_{n,m}(\varepsilon) \subseteq E_{n,m}(2\varepsilon),$$

d'où

$$\lim_{n \to \infty} \left[\lim_{m \to \infty} \Pr E_{n,m}(\varepsilon) \right] \leqq \lim_{n \to \infty} \left[\lim_{m \to \infty} \Pr F_{n,m}(\varepsilon) \right] \leqq \lim_{n \to \infty} \left[\lim_{m \to \infty} \Pr E_{n,m}(2\varepsilon) \right]$$

ce qui donne la formule de Kolmogoroff (2, note p. 315)

$$P = \lim_{\varepsilon \to 0} \left\{ \lim_{n \to \infty} \left[\lim_{m \to \infty} \Pr F_{n,m}(\varepsilon) \right] \right\} \tag{178}$$

En appelant et posant $Q = 1 - P$ et

$$Q_{n,m}(\varepsilon) = 1 - \Pr F_{n,m}(\varepsilon),$$

on peut aussi écrire

$$Q = \lim_{\varepsilon \to 0} \left\{ \lim_{n \to \infty} \left[\lim_{m \to \infty} Q_{n,m}(\varepsilon) \right] \right\} \tag{179}$$

ou encore

(180) $\Pr \{ \text{divergence de la suite des } X_n \}$

$$= \lim_{\varepsilon \to 0} \left[\lim_{n \to \infty} \left[\lim_{m \to \infty} \Pr \left\{ |X_{n+1} - X_n| \geqq \varepsilon \text{ ou } |X_{n+2} - X_n| \geqq \varepsilon \text{ ou } \ldots |X_m - X_n| \geqq \varepsilon \right\} \right] \right]$$

Il est à noter que, dans tous les cas, les limites successives figurant dans cette formule existent. De sorte que la question de l'existence

de la probabilité de la convergence d'une suite de variables aléatoires posée page 215, se trouve résolue par l'affirmative.

On conclut aussi de la formule (179) que si, pour tout $\varepsilon > 0$, $Q_{n,m}(\varepsilon)$ reste inférieur quel que soit m à une quantité $\alpha_n(\varepsilon)$ infiniment petite avec $\frac{1}{n}$, alors $Q = 0$; on procéderait de même avec P

Application. — L'application de cette formule au cas où la suite des X_n est la suite des sommes partielles $Y_1 + \dots + Y_n$ d'une série ΣY_n de variables indépendantes Y_n a fourni à M. Paul Lévy (3, p. 124) une démonstration nouvelle et simple d'une propriété résultant aussi d'un critère antérieurement démontré par MM. Khintchine et Kolmogoroff (I).

Soit en effet $G_{n,m}(\varepsilon)$ l'événement résultant pour $n < r < m$ du concours des événements $E_{n,r}(\varepsilon)$ et $E_{r,m}(\varepsilon)$. Si les Y sont indépendants, ces deux derniers événements qui ne concernent pas les mêmes Y sont indépendants et l'on a

$$\Pr. G_{n,m}(\varepsilon) = [\Pr. E_{n,r}(\varepsilon)][\Pr. E_{r,m}(\varepsilon)].$$

Or il est clair que

$$E_{n,m}(\varepsilon) \subset G_{n,m}(\varepsilon).$$

On aura donc, pour tout $\varepsilon > 0$,

$$\lim_{m \to \infty} \Pr. E_{n,m}(\varepsilon) \leqq [\Pr. E_{n,r}(\varepsilon)]\left[\lim_{m \to \infty} \Pr. E_{r,m}(\varepsilon)\right].$$

En faisant tendre dans ces inégalités, d'abord r vers l'infini, puis n vers l'infini, puis enfin ε vers zéro et sachant que toutes ces limites existent, on aura

$$P \leqq P^2, \qquad \text{d'où} \qquad P(1 - P) \leqq 0,$$

c'est-à-dire $P = 0$ ou 1. Ainsi : *la probabilité de la convergence d'une* SÉRIE *de variables aléatoires* INDÉPENDANTES *ne peut être que* 0 *ou* 1 [1].

---

[1] La proposition précédente peut aussi s'obtenir comme cas particulier d'une proposition plus générale due à M. Kolmogoroff (4, p. 60) et étendue ensuite à des cas plus généraux par MM. Jessen (4) et Paul Lévy (1, § 14). La signification intuitive de cette proposition peut s'exprimer ainsi : lorsque la probabilité d'un événement E déterminé dont la réalisation dépend des résultats d'une suite infinie d'épreuves, n'est pas modifiée par la connaissance des résultats d'un nombre fini d'épreuves, cette probabilité ne peut être égale qu'à 0 ou 1. Nous indiquerons plus loin (p. 267) l'énoncé précis de M. Kolmogoroff et sa généralisation par M. Paul Lévy.

Observons que si les Y_n sont indépendants, il en est de même de leurs valeurs absolues, donc : *le même théorème s'applique à la convergence absolue.*

Il faut aussi noter que la formule (178) de M. Kolmogoroff et le théorème ci-dessus de M. Paul Lévy s'étendent, sans changement de la démonstration, au cas où la suite des X_n ou la série des Y_n est à termes complexes.

Nous avons un peu simplifié la démonstration de M. Paul Lévy, qui faisait intervenir inutilement l'inégalité

$$\Pr G_{n,m}(\varepsilon) \leq \Pr E_{n,m}(2\varepsilon).$$

Cette simplification prend son importance quand on veut appliquer le même mode de démonstration au cas où les X_n au lieu d'être des sommes de variables indépendantes sont elles-mêmes des variables indépendantes. Dans ce cas, on peut appliquer le même raisonnement que plus haut en remplaçant toutefois $E_{r,m}$ par $E_{r+1,m}$ et supposant $n < r < m - 1$.

On a alors ce nouveau théorème qui n'est ni un cas particulier ni une généralisation du précédent :

La probabilité de la convergence d'une SUITE *de variables aléatoires* INDÉPENDANTES $X_1, X_2, \ldots, X_n$ *ne peut être que* 0 *ou* 1 (1).

Ainsi, dans deux cas importants : suite ou série de variables aléatoires *indépendantes*, s'il n'y a pas convergence presque certaine, il y a divergence presque certaine.

Cas intermédiaire. — Il ne faudrait cependant pas croire qu'on a obtenu une division ultime en deux cas seulement. Considérons en effet le cas d'une suite $X_1, X_2, \ldots$ de variables aléatoires indépendantes qui diverge presque certainement : ce cas est compatible avec la convergence en probabilité de la suite. Nous avons déjà donné un exemple de cette compatibilité; mais dans celui-ci les X_n étaient en dépendance mutuelle. Il suffit de le modifier légèrement pour avoir un exemple où les X_n sont indépendantes. Considérons en effet les mêmes fonctions $f_n(x)$ utilisées page 203, fonctions dont la suite

(1) *Voir* note de la page précédente.

diverge pour chaque valeur de x mais converge « en mesure » vers zéro. D'autre part, considérons des variables aléatoires *indépendantes* $Z_1, Z_2, \ldots, Z_n, \ldots$ restant comprises entre 0 et 1 et dont chacune a une loi de probabilité uniforme.

Enfin, prenons, pour X_n, $X_n = f_n(Z_n)$. Il est clair que les X_n convergent en probabilité vers zéro. Car, comme à la page 203,

$$\Pr\{|X_n| \geqq 0\} \leqq \Pr\{|X_n| > 0\} = \frac{1}{q}$$

et q croît indéfiniment avec n.

Or, soit $\frac{p}{q}$ le nombre rationnel de rang n, on a

$$\Pr\left\{\frac{p-\frac{1}{4}}{q} \leqq Z_n \leqq \frac{p+\frac{1}{4}}{q}\right\} = \frac{1}{2q}.$$

Et, pour

$$\frac{p-\frac{1}{4}}{q} \leqq Z_n \leqq \frac{p+\frac{1}{4}}{q},$$

on a

$$X_n \geqq \frac{3}{4}$$

et réciproquement. Donc

$$\Pr.\left\{X_n \geqq \frac{3}{4}\right\} = \frac{1}{2q}.$$

Les événements $X_n \geqq \frac{3}{4}$ sont indépendants. D'après le théorème de M. Borel, page 27, la probabilité qu'une infinité de ces événements se réalise est égale à 1, car la série des probabilités de ces événements est divergente, étant la série

$$\frac{1}{2}\left[\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+\ldots+\frac{1}{q}+\frac{1}{q}+\ldots+\frac{1}{q}+\ldots\right]$$

Il y a donc une probabilité égale à l'unité qu'une infinité des X_n soit $\geqq \frac{3}{4}$.

Le même raisonnement prouverait qu'il y a une probabilité égale à l'unité qu'une infinité des X_n soit $\leqq \frac{1}{4}$. Il y a donc aussi une probabilité égale à l'unité que ces deux événements se produisent en même temps. Or dans ce cas il y a divergence. Ainsi nous avons *un*

exemple d'une suite de variables aléatoires indépendantes qui converge en probabilité et dont pourtant la probabilité de diverger est égale à l'unité.

Nature de la limite. — Considérons d'abord une *suite* de variables aléatoires indépendantes X_n, et un intervalle numérique arbitraire (α, β). Les événements E_n consistant en ce que $X_n < \alpha$ (ou $X_n > \beta$) sont indépendants. En vertu du théorème de M. Borel, il y a donc une probabilité égale à 0 ou 1 que les E_n se produisent une infinité de fois. Il y a donc une probabilité égale à 1 ou à zéro que l'on ait $\alpha \leqq X_n \leqq \beta$ à la fois pour toutes les valeurs de n suffisamment grandes.

I. Il peut arriver que cette probabilité soit égale à zéro quels que soient α, β. Dans ce cas, la probabilité de la convergence des X_n est nulle. Sans quoi X_n convergerait avec une probabilité positive vers une variable aléatoire X. On pourrait prendre α', β' de sorte que l'événement ($X < \alpha'$ ou $X > \beta'$) ait une probabilité inférieure à l'unité Alors, en prenant $\alpha < \alpha'$, $\beta > \beta'$, la probabilité pour que $\alpha < X_n < \beta$ pour n assez grand, serait positive.

II. Supposons maintenant qu'il y ait au moins un intervalle I dans lequel soient compris les X_n à partir de n assez grand avec une probabilité égale à 1. On voit que cette propriété de I appartiendra sûrement à l'une, I_1, des deux moitiés de I, à l'une I_2 des deux moitiés de I_1, ... Soit a le nombre unique appartenant à tous les I_n. I_n étant de longueur $\frac{1}{2^n}$, il y a une probabilité égale à l'unité que les X_n soient, à partir d'un rang fini, tous compris dans l'intervalle $a - \frac{1}{2^n}$, $a + \frac{1}{2^n}$. C'est dire que la suite des X_n converge presque certainement vers le nombre a.

Le résultat obtenu est double (Paul Lévy, 3). On a d'abord obtenu une nouvelle démonstration du fait que la probabilité de la convergence des X_n est zéro ou un. Mais on a de plus démontré que : si des variables aléatoires *indépendantes* X_n ne divergent pas presque certainement, elles convergent presque certainement *vers un nombre certain.*

Ajoutons que dans le cas où X_n diverge presque certainement mais (ce qui n'est pas contradictoire) converge « en probabilité » vers une limite X, cette limite est encore un nombre certain. Car on

sait (p. 233) qu'on peut extraire des X_n une suite qui converge presque certainement vers X.

Ces compléments d'information ne sont plus admissibles quand il s'agit non d'une suite mais d'une *série*. Si parmi les séries de variables aléatoires indépendantes convergeant presque certainement, il peut en exister (soit, par exemple, $X_1 + X_2 + \dots$) dont la somme est un nombre certain (par exemple A), il ne peut en être de même pour toutes ces séries. Il suffit, en effet, pour le voir, de considérer une variable aléatoire X_0, non constante et indépendante de $X_1, X_2, \dots$, et de former, par exemple, la série $X_0 + X_1 + X_2 + \dots$ qui converge presque certainement vers le nombre non certain $A + X_0$.

Suites de sommes. — Observons que si deux suites de variables aléatoires $X_1, X_2, \dots, X''_1, X''_2, \dots$ convergent presque certainement, il en est de même de la suite de leurs sommes $X_1 + X''_1, X_2 + X''_2, \dots, X_n + X''_n, \dots$ En effet, la probabilité Q de la divergence de l'une au moins des séries ΣX_n, $\Sigma X''_n$ est au plus égale à la somme des probabilités de ces divergences, c'est-à-dire à zéro. Or la probabilité de la divergence de $\Sigma(X_n + X''_n)$ est au plus égale à Q.

De plus, si X, X'' sont les limites presque certaines de X_n, X''_n, alors, en dehors d'un événement de probabilité $Q = 0$, X_n et X''_n convergent au sens ordinaire vers $X + X''$, donc $X_n + X''_n$ vers $X + X''$; par suite $X + X''$ est la limite presque certaine de $X_n + X''_n$.

Suites comparables. — Il en résulte que, pour établir la convergence presque certaine de X_n, il pourra y avoir avantage à former, et cela suffira, une suite X'_n qui converge presque certainement de sorte que la suite $X_n - X'_n$ converge aussi presque certainement. Et si l'on peut choisir X'_n de sorte que zéro soit la limite presque certaine de $X_n - X''_n$, les limites presque certaines de X'_n et X_n seront les mêmes.

Prenons maintenant plus généralement pour les X_n une suite quelconque et les X'_n tels que $X'_n - X_n$ converge presque certainement vers zéro. Si l'on connaît la valeur p de la probabilité de la convergence de X_n, la valeur p' de la probabilité de X'_n sera, sous ces conditions, la même. Car si E est l'événement consistant dans la convergence de X_n, F celui consistant dans la convergence de $X'_n - X_n$ vers zéro et E_1 celui consistant dans le concours de E et de F, on a (p. 28), puisque $\text{Pr. } F = 1$, $p = \text{Pr. } E_1$. Or $X'_n = X_n + (X'_n - X_n)$, donc,

quand E_1 a lieu X'_n converge. Par suite Pr. E_1 est au plus égale à la probabilité p' de la convergence de X'_n. Ainsi $p \leqq p'$. De même $p' \leqq p$ D'où $p = p'$. D'ailleurs, si p_0 est la probabilité de la convergence simultanée de X_n et X'_n, on a $p(E_1) \leqq p_0 \leqq p$, d'où $p_0 = p$ Ainsi $p = p' = p_0$. De plus en cas de convergence simultanée de X_n et X'_n, leurs limites sont presque certainement égales.

Supposons encore que $X_n - X'_n$ converge presque certainement vers zéro, alors les probabilités respectives ϖ, ϖ' de la convergence *en moyenne arithmétique* de la suite des X_n et de celle des X'_n sont égales entre elles, elles sont égales à la probabilité ϖ_0 de la convergence en moyenne simultanée des X_n et des X'_n et, dans ce dernier cas, les sommes généralisées de X_n et X'_n sont presque certainement égales. Et même il suffit de supposer que $X_n - X'_n$ converge en moyenne presque certainement vers zéro

En effet, ϖ, ϖ' sont les probabilités des convergences au sens ordinaire des variables aléatoires $S_n = \dfrac{X_1 + \quad + X_n}{n}$, $S'_n = \dfrac{X'_1 + \quad + X'_n}{n}$ et dans l'hypothèse faite $S_n - S'_n$ converge presque certainement vers zéro.

Suites équivalentes. — Ce sont sans doute d'abord ces considérations qui ont incité M. Khintchine à formuler la définition des *suites équivalentes*. Deux suites de variables aléatoires Y_n, Y'_n seront dites équivalentes si la série

$$\sum_n \text{Pr} \{ Y_n \neq Y'_n \}$$

est convergente. Cette série majore, quel que soit $\varepsilon > 0$, la série

$$\sum_n \text{Pr.} \{ | Y_n - Y'_n | > \varepsilon \},$$

donc (p. 242) $Y_n - Y'_n$ converge presque certainement vers zéro. Il en résulte que si la suite des Y'_n converge presque certainement, il en sera de même de la suite des Y_n et avec la même limite.

A cette même propriété (qui s'appliquerait encore si l'on savait seulement que $Y_n - Y'_n$ converge presque certainement vers zéro) s'ajoute une propriété spéciale aux suites équivalentes. C'est que l'événement consistant en ce que les deux suites Y_n et Y'_n ont les

mêmes termes à partir d'un certain rang (non donné d'avance) a une probabilité égale à l'unité. (Car d'après le théorème de la page 26, l'événement consistant dans la réalisation d'une infinité des événements $Y_n \neq Y'_n$ aura une probabilité nulle.) Il en résulte qu'on pourra souvent substituer utilement l'une des suites à l'autre dans une proposition où l'on néglige les événements de probabilités nulles et qui n'est pas affectée par le changement d'un nombre fini de termes (ce nombre de termes n'étant pas fixé d'avance).

En particulier, considérons les deux séries

$$Y_1 + Y_2 + \ldots \qquad Y_1 + Y'_2 + \ldots$$

et soient P, P' les probabilités des événements E, E' consistant dans leurs convergences respectives. Soit E_1 l'événement consistant dans le concours de E et de l'événement e consistant dans le fait que les deux suites ont les mêmes termes à partir d'un rang fini (arbitraire). Quand E_1 a lieu, $\Sigma Y'_n$ converge donc $\text{Pr.}\, E_1 \leqq P'$. Mais, puisque $\text{Pr.}\, e = 1$, on a

$$\text{Pr.}\, E_1 = \text{Pr.}\, E = P.$$

D'où $P \leqq P'$ et de même $P' \leqq P$. On observe en outre que si E_1 a lieu, les deux séries sont convergentes en même temps. Si P_0 est la probabilité de la convergence simultanée des deux séries, on a donc $\text{Pr.}\, E_1 \leqq P_0$; d'où $P \leqq P_0$. Or, même si les deux suites Y_n, Y'_n n'étaient pas équivalentes, on aurait évidemment $P_0 \leqq P$. Donc $P = P_0 = P'$.

Ainsi : *quand les termes de deux séries ΣY_n, $\Sigma Y'_n$ forment deux suites équivalentes (au sens de Khintchine) les probabilités de leurs convergences respectives sont égales et égales à celle de leur convergence simultanée.* Il est bien entendu que, dans ce dernier cas, les sommes des deux séries ne sont pas, en général, égales.

Un des avantages de la notion introduite est qu'on peut toujours, étant donnée une suite X_n, déterminer une suite équivalente X'_n, dont chacun des termes est borné. Soit en effet $\Sigma \omega_n$ une série convergente à termes positifs certains. Il existe un nombre $A_n > 0$, tel que

$$\text{Pr}\, \{ |X_n| > A_n \} < \omega_n.$$

Il suffit alors de prendre

$$X'_n = X_n \qquad \text{pour } |X_n| \leqq A_n,$$
$$X'_n = 0, \qquad \text{pour } |X_n| > A_n.$$

La série

$$\Sigma \Pr\{X'_n \neq X_n\} = \Sigma \Pr\{|X_n| > A_n\}$$

étant majorée par $\Sigma \omega_n$, sera bien convergente.

— Nous allons maintenant indiquer des formes de critères plus commodes pour la convergence presque certaine. Ceux-ci feront, en effet, intervenir la convergence de séries formées au moyen de valeurs moyennes ou d'écarts quadratiques moyens, nombres généralement plus faciles à calculer que les termes de (176). Par contre, ils concernent des suites moins générales que les précédentes.

Critère de Kolmogoroff. — Considérons à nouveau le cas où Z_1, ..., Z_n, ... sont des variables aléatoires *indépendantes*, de moyennes nulles : $\overline{Z_k} = 0$ et où l'on a posé

$$(\mu_k)^2 = \mathfrak{M}(Z_k)^2 \qquad S_k = Z_1 + \ldots + Z_k.$$

On se propose d'étudier *la convergence de*

$$v_n = \frac{S_n}{n} = \frac{1}{n}(Z_1 + \ldots + Z_n).$$

Pour tout entier n, il existe un entier $m \geqq 0$ tel que $2^m \leqq n < 2^{m+1}$, soit alors P_m la probabilité pour qu'ait lieu l'inégalité $|v_n| > \varepsilon$ pour au moins une des valeurs de n telles que $2^m \leqq n < 2^{m+1}$. Cette probabilité est au plus égale à la probabilité qu'ait lieu $|S_n| > 2^m \varepsilon$ pour l'une au moins des mêmes valeurs de n. Dès lors, d'après l'inégalité (100) de Kolmogoroff (p. 130), on a ici

$$P_m \leqq \frac{1}{(2^m \varepsilon)^2} \sum_{j < 2^{m+1}} \mu_j^2.$$

Si maintenant, on appelle P'_ν la probabilité pour que l'on ait $|v_n| < \varepsilon$ pour au moins une valeur entière de $n \geqq \nu$, on aura

$$P'_\nu \leqq \sum_{m=\rho}^{m=+\infty} P_m \leqq \frac{1}{\varepsilon^2} \sum_{m=\rho}^{m=+\infty} \frac{1}{2^{2m}} \sum_{j<2^{m+1}} \mu_j^2,$$

où $2^\rho \leq \nu < 2^{\rho+1}$. D'où

$$P'_\nu \leq \frac{1}{\varepsilon^2}\sum_{r \geq \rho}\left[\sum_{m \geq r}\frac{1}{2^{2m}}\right]\left[\sum_{2^r \leq j < 2^{r+1}} \mu_j^2\right]$$
$$= \frac{1}{\varepsilon^2}\sum_{r \geq \rho}\frac{16}{3}\,\frac{1}{2^{2(r+1)}}\sum_{2^r \leq j < 2^{r+1}} \mu_j^2 \leq \frac{16}{3\varepsilon^2}\sum_{j \geq 2^\rho}\frac{\mu_j^2}{j^2}.$$

Il suffit pour que ces calculs soient exacts que la série

$$\sigma = \sum_n \frac{\mu_n^2}{n^2}$$

soit convergente. En appelant $R(n)$ son reste de rang n, on aura

$$P'_\nu = \frac{16}{3\varepsilon^2} R(2^\rho)$$

Quand $\nu \to \infty$, ρ croît aussi indéfiniment et $R(2^\rho)$ tend vers zero. Dès lors, quel que soit $\varepsilon > 0$, P'_ν tend vers zéro avec $\frac{1}{\nu}$; v_n converge vers zéro fortement et, par suite, presque certainement.

Nous avons vu, page 239, que cette convergence presque certaine a lieu quand les Z_n sont bornés ou plus généralement quand leurs écarts moyens d'ordre 4 sont bornés. Nous voyons maintenant qu'il en sera encore ainsi dans le cas à la fois plus général et plus simple *où les écarts quadratiques moyens des Z_n sont bornés dans leur ensemble.* Ce résultat s'obtenait aussi comme cas particulier d'une première généralisation due à M. Khintchine (2). Le résultat obtenu maintenant est plus général encore (Kolmogoroff, 3) :

Si $Z_1, Z_2, \ldots$ est une suite de variables aléatoires indépendantes dont les valeurs moyennes $\bar{Z}_n$ sont nulles, la suite de leurs moyennes arithmétiques $v_n = \frac{Z_1 + \ldots + Z_n}{n}$ converge presque certainement vers zéro quand la série $\sum_n \mathfrak{M}\left(\frac{Z_n}{n}\right)^2$ converge.

Il en résulte l'extension suivante du théorème de Cantelli de la p. 239 :

Soient $Y_1, Y_2, \ldots$ des variables aléatoires indépendantes dont les valeurs moyennes $\bar{Y}_1, \bar{Y}_2, \ldots$ tendent vers une limite V. Pour que

la suite de leurs moyennes arithmétiques v_n converge uniformément « en probabilité » vers V, il suffit que les écarts quadratiques moyens des Y_n soient bornés dans leur ensemble ou plus généralement que la série $\sum_n \mathfrak{M}\left(\frac{Y_n}{n}\right)^2$ soit convergente.

— En particulier, considérons *une* variable aléatoire X dont les valeurs $X_1, X_2, \ldots, X_n$, au cours de n épreuves sont supposées indépendantes. Alors en posant $Z_n = X_n - \overline{X}$, les Z_n sont indépendants et les $\overline{Z_n}$ sont nuls. Si l'écart quadratique moyen μ de X est fini, la la série $\sum \mathfrak{M}\left(\frac{Z_n}{n}\right)^2$ se réduit à

$$\sum \frac{\mu^2}{n^2} = \mu^2 \sum \frac{1}{n^2};$$

elle est donc convergente Ainsi sous les seules hypothèses que X et μ^2 sont déterminés et finis, il est démontré que $\frac{X_1 + \ldots + X_n}{n}$ converge presque certainement vers $\overline{X}$. Ce résultat, obtenu directement par M. Khintchine (2), a été étendu par M. Kolmogoroff (*voir* plus loin, p. 257) au cas où μ n'est pas supposé fini

Revenons au critère général

On peut dire, en un certain sens, qu'on ne peut donner un théorème plus précis que le précédent. En effet, si l'on considère réciproquement une suite de nombres $\mu'_n \geqq 0$ tels que la série

$$\sum_n \left(\frac{\mu'_n}{n}\right)^2$$

diverge, alors il existe une suite de variables aléatoires indépendantes Z_n, telles que

$$(184) \qquad \overline{Z_n} = 0, \qquad \mathfrak{M}(Z_n)^2 = (\mu'_n)^2$$

et que $v_n = \frac{Z_1 + \ldots + Z_n}{n}$ ne converge pas presque certainement vers zéro. Il suffit de prendre avec M. Kolmogoroff (3), les Z_n de la façon suivante :

1. Si $\mu'_n \leqq n$, $Z_n = (n, -n$ ou $0)$ avec les probabilités respectives $\frac{\mu'^2_n}{2n^2}, \frac{\mu'^2_n}{2n^2}, 1 - \frac{\mu'^2_n}{n^2}$.

II. Si $\mu'_n > n$, $Z_n = (\mu'_n$ ou $-\mu'_n)$ avec les probabilités respectives $\frac{1}{2}$ et $\frac{1}{2}$.

On vérifie immédiatement que les Z_n sont indépendants et satisfont aux conditions (184).

D'autre part, comme

$$\frac{Z_n}{n} = v_n - \left(1 - \frac{1}{n}\right) v_{n-1},$$

$\frac{Z_n}{n}$ converge vers zero quand v_n tend vers zéro ; donc si v_n convergeait presque certainement vers zéro, il devrait en être de même de $\frac{Z_n}{n}$. Or, si $\frac{Z_n}{n} \neq 0$, on a, dans tous les cas,

$$\left|\frac{Z_n}{n}\right| \geqq 1$$

et cet événement a la probabilité $\frac{\mu'^2_n}{n^2}$, si $\mu'_n \leq n$ ou 1 si $\mu'_n \geq n$. Donc pour tout ε entre 0 et 1, on a

$$\Pr\left\{\left|\frac{Z_n}{n}\right| > \varepsilon\right\} = \left(\frac{\mu'^2_n}{n^2} \text{ ou } 1\right).$$

Or les $\frac{Z_n}{n}$ étant indépendants, pour qu'ils convergent presque certainement vers zéro, il faudrait (p. 243) que la série

$$\sum_n \Pr\left\{\left|\frac{Z_n}{n}\right| > \varepsilon\right\}$$

fût convergente. Dans ce cas, les termes égaux à 1 de cette série seraient en nombre fini. Et à partir du dernier, de rang $r-1$, les termes de cette série seraient les mêmes que ceux de $\sum_{n=r}^{n=+\infty} \frac{\mu'^2_n}{n^2}$ qui, par hypothèse, n'est pas convergente.

Si l'on convient de dire qu'une suite $u_1, u_2, \ldots$ converge *au sens de Cesaro* vers a, quand la suite des moyennes arithmétiques $\frac{u_1 + \ldots + u_n}{n}$ converge vers a, on voit que le théorème de Kolmogoroff fournit un critère de convergence au sens de Cesaro,

presque certaine, vers zéro d'une suite de variables aléatoires indépendantes Z_n dont les valeurs moyennes $\overline{Z}_n$ sont nulles.

Généralisation de Kolmogoroff. — La proposition de la page 255 a été étendue par M. Kolmogoroff au cas où l'on supprime toute hypothèse sur l'existence d'un écart quadratique moyen. M. Kolmogoroff a bien voulu nous communiquer sa démonstration (encore inédite) de ce théorème.

Soit donc X une variable aléatoire qui n'est soumise qu'à la seule condition d'avoir une vraie valeur moyenne. C'est-à-dire que cette valeur moyenne

$$\overline{X} = \int_{-\infty}^{+\infty} x\, dF(x)$$

est non seulement la limite finie de $\int_{-\alpha}^{+\alpha} x\, dF(x)$ quand $\alpha \to +\infty$, mais aussi la limite finie et unique de $\int_{\beta}^{\alpha} x\, dF(x)$ quand β et α tendent simultanément mais indépendamment vers $-\infty$ et $+\infty$ respectivement.

Il en résulte que si l'on pose

$$q(t) = \Pr.\ \{|X| < t\},$$

l'intégrale

$$W = \int_{0}^{+\infty} t\, dq(t),$$

qui représente la valeur moyenne de $|X|$, est aussi finie.

Soient maintenant $X_1, X_2, \ldots, X_n, \ldots$ les valeurs prises par X dans une suite d'épreuves indépendantes. Posons

$$\begin{aligned} Y_n &= X_n \qquad \text{pour } |X_n| < n, \\ Y_n &= 0 \qquad \text{pour } |X_n| \geqq n. \end{aligned}$$

Nous démontrerons le théorème pour les X_n en démontrant que les X_n et les Y_n forment des suites équivalentes au sens de M. Khintchine (p. 251) et en prouvant le théorème pour les Y_n.

Il s'agit d'abord de prouver que la série $\Sigma \varepsilon_n$ est convergente en posant

$$\begin{aligned} \varepsilon_n &= \Pr.\ \{X_n \neq Y_n\} = \Pr.\ \{|X_n| \geqq n\} = 1 - q(n) \\ &= [q(n+1) - q(n)] + [q(n+2) - q(n+1)] + \ldots . \end{aligned}$$

D'où

$$\sum_{n=1}^{n=+\infty} \varepsilon_n = \sum_{n=1}^{n=+\infty} n[q(n+1)-q(n)] \leqq \int_0^{+\infty} t\,dq(t) = W$$

Ainsi $\Sigma\varepsilon_n$ est convergente et les suites des X_n et des Y_n sont équivalentes.

Considérons donc la suite des Y_n. On a vu (p. 254) que si μ_n est l'écart quadratique moyen de Y_n, alors la convergence de $\sum \frac{\mu_n^2}{n^2}$ suffit pour que $\frac{Y_1+\dots+Y_n}{n} - \frac{\overline{Y}_1+\dots+\overline{Y}_n}{n}$ converge presque sûrement vers zéro. On a

$$\mu_n^2 = \mathfrak{M}(Y_n - \overline{Y}_n)^2 \leqq \mathfrak{M} Y_n^2 = \int_0^{+\infty} t^2\,dq_n(t),$$

où

$$q_n(t) = \Pr\{|Y_n| < t\}.$$

On a

$$|Y_n| < n, \qquad \text{d'où} \qquad q_n(t) = 1 \qquad \text{pour } t > n$$

Par contre pour $0 < t < n$, on a

$$q_n(t) = q(t) + 1 - q(n).$$

Donc

$$\int_0^{+\infty} t^2\,dq_n(t) = \int_0^n t^2\,dq(t)$$

et, par suite,

$$\mu_n^2 \leqq \int_0^n t^2\,dq(t)$$

La série

$$\sum_{n=1}^{n=+\infty} \frac{\mu_n^2}{n^2}$$

est donc majorée par la série

$$\sum_{n=1}^{n=+\infty} \frac{1}{n^2}\int_0^n t^2\,dq(t) = \sum_{n=1}^{n=+\infty} \left\{\left[\int_{n-1}^n t^2\,dq(t)\right]\left[\sum_{r \geqq n} \frac{1}{r^2}\right]\right\}.$$

Or

$$\sum_{r \geqq n} \frac{1}{r^2} \leqq \sum_{r \geqq n} \frac{1}{(r-1)r} = \sum_{r \geqq n}\left[\frac{1}{r-1} - \frac{1}{r}\right] = \frac{1}{n-1} \leqq \frac{2}{n} \qquad \text{pour } n \geqq 2$$

Donc $\sum_{n\geq 2}\frac{\mu_n^2}{n^2}$ est majorée par la série

$$2\sum_{n\geq 2}\int_{n-1}^{n}\frac{t^2}{n}\,dq(t)\leq 2\sum_{n\geq 2}\int_{n-1}^{n}t\,dq(t)\leq 2\mathrm{W}$$

Ainsi la série $\sum\frac{\mu_n^2}{n^2}$ est convergente. Les Y_n sont indépendants comme les X_n, par suite $\frac{\mathrm{Y}_1+\ldots+\mathrm{Y}_n}{n}-\frac{\overline{\mathrm{Y}}_1+\ldots+\overline{\mathrm{Y}}_n}{n}$ converge presque sûrement vers zéro.

D'ailleurs $\frac{\overline{\mathrm{Y}}_1+\ldots+\overline{\mathrm{Y}}_n}{n}$ tend vers $\overline{\mathrm{X}}=\mathrm{V}$. Il suffit, pour le voir, de montrer que $\overline{\mathrm{Y}}_n$ tend vers $\overline{\mathrm{X}}$, on sait déjà que $\mathrm{Y}_n-\mathrm{X}_n$ converge presque certainement vers zéro. De plus, si

$$\mathrm{F}_n(x)=\mathrm{Pr}\,\{\mathrm{Y}_n<x\},$$

on a

$$\mathrm{F}_n(x)=0\qquad\text{si } x\leq -n,$$
$$\mathrm{F}_n(x)=1\qquad\text{si } x\geq n$$

Si $-n<x\leq 0$, on a

$$\mathrm{F}_n(x)=\mathrm{F}(x)-\mathrm{F}(-n+0)$$

Si $0<x<n$,

$$\mathrm{F}_n(x)=\mathrm{F}(x)+1-\mathrm{F}(n)$$

Donc

$$\overline{\mathrm{Y}}_n=\int_{-\infty}^{+\infty}x\,d\mathrm{F}_n(x)=\int_{-n}^{+n}x\,d\mathrm{F}_n(x)=\int_{-n}^{+n}x\,d\mathrm{F}(x)\quad(^1)$$

et, par suite, $\overline{\mathrm{Y}}_n$ tend bien vers $\overline{\mathrm{X}}$.

Donc $\frac{\mathrm{Y}_1+\ldots+\mathrm{Y}_n}{n}$ converge presque sûrement vers $\overline{\mathrm{X}}$ et puisque les Y_n et les X_n sont deux suites équivalentes, il en est de même de $\frac{\mathrm{X}_1+\ldots+\mathrm{X}_n}{n}$.

En résumé : *Soit* X *une variable aléatoire dont on suppose seulement que* X *et* $|\mathrm{X}|$ *ont chacune une valeur moyenne finie, soient* X_1, X_2, ... *les valeurs prises par* X *dans une suite d'épreuves indépendantes; alors* X_n *converge au sens de Cesaro presque certainement vers la valeur moyenne* $\overline{\mathrm{X}}$ *de* X.

---

(1) La dernière égalité suppose F continue pour $x=-n$. Dans le cas contraire, on écrirait que φ_n est compris entre $\int_{-n+\theta_n}^{n}x\,d\mathrm{F}$ et $\int_{-n-\theta_n}^{n}x\,d\mathrm{F}$ où $0<\theta_n<1$

Théorème de Glivenko-Cantelli. — A titre d'application du théorème de Polya-Cantelli (p. 34), considérons une variable aléatoire X et $X_1, \ldots, X_n$ ses valeurs dans n épreuves indépendantes. Soit E_x l'événement consistant en ce que $X < x$, la fonction des probabilités totales de X est la probabilité $F(x)$ de E_x, la valeur empirique de $F(x)$ est la fréquence $F_n(x) = \frac{R_n(x)}{n}$ de E_x dans les n épreuves. D'après le théoreme de Borel, on sait que $F_n(x)$, qui est une variable aléatoire, converge presque certainement quand n tend vers l'infini, vers $F(x)$, qui est un nombre certain pour x arbitraire, mais fixé. Dans le cas actuel, cet énoncé peut être précisé.

Les nombres $F(x)$ et $F_n(x)$ étant compris entre 0 et 1, la valeur absolue de leur différence a, pour n fixe, une borne supérieure finie L_n, quand x varie de $-\infty$ à $+\infty$. L_n est aussi une variable aléatoire ; que devient-elle quand n croît ? M. Glivenko (2) a posé la question et a démontré que si $F(x)$ est continue, L_n tend presque certainement vers zéro. M. Cantelli (9, 10) a étendu ce résultat au cas des fonctions $F(x)$ non nécessairement continues.

On peut démontrer ce résultat de la manière suivante Soient d_1, d_2, ... les points de discontinuité, s'il en existe de $F(x)$. La probabilité que $X = d_i$ est égale à $F(d_i + 0) - F(d_i)$, c'est presque certainement la limite de la fréquence $F_n(d_i + 0) - F_n(d_i)$ avec laquelle la valeur d_i est atteinte dans les n premières épreuves Soit maintenant $c_1, c_2, \ldots$ une suite formée, dans un ordre quelconque, des nombres rationnels (>0, <0 ou 0) et des abscisses d_1, d_2, ... On a vu que $F_n(c_i)$ converge presque certainement vers $F(c_i)$. Il y a une probabilité nulle pour chacun des contraires des événements des deux classes qu'on vient de considérer. Comme l'ensemble de ces événements est dénombrable, il y a une probabilité nulle pour que l'un au moins de ces contraires ait lieu (p. 24).

Dès lors, sauf quand un événement E de probabilité nulle a lieu, toutes les quantités $F_n(d_i + 0) - F_n(d_i)$ tendent vers les quantités $F(d_i + 0) - F(d_i)$, et toutes les quantités $F_n(c_i)$ tendent vers les $F(c_i)$.

En particulier, les $F_n(d_j)$ tendent alors vers les $F(d_j)$ et, par suite, les $F_n(d_j + 0)$ vers les $F(d_j + 0)$. Comme les F_n et F sont continues hors des points d_j et qu'elles sont partout continues à gauche, on voit que $F_n(c_j)$, $F_n(c_j + 0)$, $F_n(c_j - 0)$ convergent

respectivement vers $F(c_i)$, $F(c_j+0)$, $F(c_j-0)$, en dehors de l'événement E. La même propriété a lieu en remplaçant c_i par un point x quelconque : il suffit de le montrer quand x est un point de continuité de $F(x)$. Tout point x peut être encadré entre un point c_k et un point c_l; $F_n(x\pm 0)$ sont compris entre $F_n(c_k)$ et $F_n(c_l)$ qui tendent vers $F(c_k)$ et $F(c_l)$. Donc les limites de $F_n(x\pm 0)$ sont comprises entre $F(c_k)$ et $F(c_l)$, lesquels peuvent être supposés aussi voisins que l'on veut de $F(x)$. Dès lors, $F_n(x\pm 0)$, et par suite, $F_n(x)$ convergent vers $F(x)$. On peut alors dire que $F_n(x+0)$, $F_n(x-0)$, $F_n(x)$ convergent respectivement vers $F(x+0)$, $F(x-0)$, $F(x)$.

En vertu du théoreme de Polya-Cantelli de la page 34, $F_n(x)$ converge alors vers $F(x)$ uniformément sur tout l'intervalle $(-\infty, +\infty)$, en dehors de l'événement E.

En résumé, sans qu'il soit utile de restreindre la généralité de la fonction des probabilités totales $F(x)$ de X, il est établi que *si $F_n(x)$ est la valeur déterminée empiriquement, au moyen de n expériences indépendantes, de la fonction des probabilités totales $F(x)$ d'une variable aléatoire X, il y a une probabilité égale à l'unité que la fonction $F_n(x)$ (qui est aléatoire) converge vers la fonction certaine $F(x)$ (quand n croît) uniformément sur tout l'intervalle* $(-\infty, +\infty)$

Convergence d'une série de variables indépendantes. — Considérons maintenant une *série* de variables aléatoires *indépendantes*

$$X_1+X_2+\ldots+X_n+\ldots \tag{185}$$

Pour étudier la convergence de cette série, il faut étudier la convergence de la suite de variables aléatoires $s_n=X_1+\ldots+X_n$. Contrairement au cas précédent on n'a donc pas ici à étudier la convergence d'une suite de variables indépendantes.

On a vu (p. 246) que ΣX_n ou bien converge presque certainement, ou bien diverge presque certainement.

MM. Khintchine et Kolmogoroff (1) ont donné dans ce cas un critère de convergence.

Une démonstration plus simple du même critère a été donnée par M. Kolmogoroff (**2**, p. 314). A cet effet, posons

$$U_n=X_n-\overline{X_n}, \qquad \mu_n^2=\mathfrak{M}(X_n-\overline{X_n})^2$$

et appelons $Q_{n,m}(\varepsilon)$ la probabilité que l'inégalité

$$|U_n + U_{n+1} + \ldots + U_r| \geq \varepsilon$$

soit vérifiée pour l'une au moins des valeurs de r de n à $m (m \geq n)$.

En vertu de l'inégalité (100) de Kolmogoroff, on aura ici

$$(186) \qquad Q_{n,m}(\varepsilon) \leq \frac{1}{\varepsilon^2}(\mu_n^2 + \ldots + \mu_m^2).$$

Or, la probabilité Q de la divergence de ΣU_n peut s'exprimer (p. 245) sous la forme

$$(179) \qquad Q = \lim_{\varepsilon \to 0} \left\{ \lim_{n \to \infty} \left[\lim_{m \to \infty} Q_{n,m}(\varepsilon) \right] \right\}.$$

La formule (186) ne peut fournir un renseignement sur $\lim_{m \to \infty} Q_{n,m}(\varepsilon)$ que si la série $\sum_n \mu_n^2$ est convergente. Faisons cette hypothèse. Alors la combinaison de (186) et (179) fournit Q = 0 de sorte que $U_1 + \ldots + U_n$ converge presque certainement vers une limite. Or

$$X_1 + \ldots + X_n = (U_1 + \ldots + U_n) + (\overline{X}_1 + \ldots + \overline{X}_n).$$

Il y aura donc convergence presque certaine de ΣX_n, si l'on fait la supposition que la série certaine $\Sigma \overline{X}_n$ converge.

En résumé : *si les variables aléatoires* X_n *sont indépendantes et si la série de leurs valeurs moyennes* $\Sigma \overline{X}_n$ *et la série des carrés de leurs écarts quadratiques moyens,* $\Sigma \mathfrak{M}(X_n - \overline{X}_n)^2$ *sont convergentes, alors la série* ΣX_n *converge presque certainement.*

Application. — Comme application de cette proposition, on peut obtenir un résultat démontré antérieurement par M. Steinhaus (1) Considérons une série à termes complexes :

$$S = \Sigma c_n e^{i\Phi_n},$$

où $c_n = a_n + ib_n$ est un nombre *certain* et où les Φ_n sont des angles aléatoires supposés *indépendants*. Nous savons déjà que : la probabilité de la convergence de S est égale à zéro ou l'unité.

Pour appliquer le critère de Khintchine et Kolmogoroff, considérons séparément les parties réelles et imaginaires de S

$$S = S' + iS''.$$

On a, par exemple,

$$S' = \Sigma X'_n$$

avec

$$X'_n = a_n \cos \Phi_n - b_n \sin \Phi_n$$

On peut écrire

$$X'_n = \rho_n \cos(\Phi_n - \alpha_n),$$

où ρ_n et α_n sont des nombres certains. On a

$$\overline{X'_n} = \rho_n \mathfrak{M} \cos(\Phi_n - \alpha_n)$$

La valeur moyenne de $\cos(\Phi_n - \alpha_n)$ est un nombre u_n tel que $|u_n| < 1$. Donc

$$\mu'^2_n = \mathfrak{M}(X'_n - \overline{X'_n})^2 = \rho_n^2 \mathfrak{M}[\cos(\Phi_n - \alpha_n) - u_n]^2 \leqq 4\rho_n^2$$

Pour que $\Sigma \mu'^2_n$ converge, il suffit donc que $\Sigma \rho_n^2 = \Sigma |c_n|^2$ converge

Quand cette condition est réalisée, $\Sigma(X'_n - \overline{X'_n})$ converge presque certainement, de même $i\Sigma(X''_n - \overline{X''_n})$ converge presque certainement. Il en est donc ainsi (p. 28) de leur somme $\Sigma(X_n - \overline{X_n})$.

Le résultat est encore plus simple dans le cas particulier considéré par M. Steinhaus celui où la loi de probabilité de chaque Φ_n est uniforme. Dans ce cas, on a, par exemple,

$$u_n = \frac{1}{2\pi} \int_0^{2\pi} \cos(\varphi - \alpha_n)\, d\varphi = 0,$$

et, par suite, $\overline{X'_n} = 0$ et, de même aussi, $\overline{X''_n} = 0$. De sorte que ΣX_n converge aussi. Alors : *si les angles Φ_n sont indépendants et suivent chacun la loi de probabilité uniforme, la convergence de la série certaine $\Sigma |c_n|^2$ suffit à assurer la convergence presque certaine de la série aléatoire $\Sigma c_n e^{i\Phi_n}$.* M. Steinhaus a montré que *cette condition est aussi nécessaire.* Il y a lieu de rappeler que si la série $\Sigma |c_n|$ est convergente (et alors $\Sigma |c_n|^2$ est aussi convergente), la série $\Sigma c_n e^{i\Phi_n}$ est certainement absolument convergente, que les Φ_n soient ou non indépendantes et quelles que soient leurs lois de probabilité.

Généralisation. — Grâce à la notion de suites équivalentes de M. Khintchine, la portée du critère de MM. Khintchine et Kolmogoroff a pu être étendue par ces auteurs.

Considérons, en effet, la série de variables aléatoires indépendantes ΣX_n : elle sera, comme on l'a vu (p. 252), encore convergente presque certainement quand il en sera ainsi pour une série équivalente et, par suite, s'il existe une série de variables aléatoires $\Sigma X'_n$ telle que : a, les séries certaines $\Sigma \overline{X'_n}$, $\Sigma \mathfrak{M}(X_n - X'_n)^2$ sont convergentes, b, les suites X_n, X'_n sont équivalentes.

Réciproque. — On a vu plus haut que si les X_n sont indépendants, la probabilité de la convergence ne peut être que 0 ou 1. MM. Khintchine et Kolmogoroff ont montré que la condition généralisée ci-dessus, suffisante pour la convergence presque certaine de ΣX_n est aussi nécessaire. D'où résulte que s'il n'existe aucune suite X'_n satisfaisant aux conditions a, b ci-dessus, la probabilité de la convergence de ΣX_n est nulle.

M. Paul Lévy (3, p. 129-131) a donné plus tard une nouvelle démonstration de cette réciproque.

SECTION VI SUITES ASYMPTOTES

Définition. — Nous généraliserons ici une heureuse conception introduite (sous un nom un peu différent) par M. Slutsky (1, p. 40). Dans de nombreuses circonstances, dont nous avons rencontré quelques exemples, on a affaire à deux suites de variables aléatoires $X_1, X_2, \ldots, Y_1, Y_2, \ldots$ dont la différence devient probablement petite en valeur absolue quand n croît indéfiniment. Nous dirons (en généralisant le point de vue de M. Slutsky lié à la convergence « en probabilité ») que ces deux suites sont asymptotes en moyenne d'ordre r, ou en probabilité, ou presque certainement, suivant que $|X_n - Y_n|$ converge vers zéro avec $\frac{1}{n}$ en moyenne d'ordre r, ou en probabilité, ou presque certainement.

Énoncés. — Cette manière de parler permet d'exprimer sous une forme plus simple et plus frappante un certain nombre de propositions importantes.

Par exemple, le théorème de Poisson de la page 40, pourra s'énoncer ainsi : si dans une suite infinie d'épreuves, un événement E a une probabilité déterminée à chaque épreuve, la suite de ses fréquences $f^{(n)}$ est presque certainement asymptote à la suite des moyennes arithmétiques $p^{(n)}$ des probabilités de E.

D'ailleurs si deux suites S′ et S″ sont asymptotes toutes deux en moyenne d'ordre r, ou toutes deux « en probabilité », ou toutes deux presque certainement, à une même suite S, alors S′ et S″ sont aussi asymptotes (respectivement en moyenne d'ordre r, ou « en probabilité » ou « presque certainement »).

Soient, en effet, X_n, X'_n, X''_n les termes généraux respectifs de S, S′, S″.

Si S′ et S″ sont asymptotes à S en moyenne d'ordre r, alors $\overline{|X'_n - X_n|^r}$ et $\overline{|X''_n - X_n|^r}$ tendent vers zéro avec $\frac{1}{n}$. Or, posons

$$u = X'_n - X_n, \qquad v = X_n - X''_n, \qquad |u|^r = \alpha \qquad |v|^r = \beta$$

On a, d'après (160ᶜ, p. 189),

$$\mathfrak{M}\,|X'_n - X''_n|^r \leqq L\,\mathfrak{M}\,|X_n - X'_n|^r + L\,\mathfrak{M}\,|X''_n - X_n|^r$$

Alors puisque α et β ont, pour n assez grand, des valeurs moyennes finies, il en sera de même de $|X'_n - X''_n|^r$, et comme les valeurs moyennes de α et de β tendent vers zéro, il en sera de même de celle de $|X'_n - X''_n|^r$. De sorte que S′ et S″ seront asymptotes en moyenne d'ordre r.

Si $X'_n - X_n$ et $X_n - X''_n$ convergent en probabilité vers zéro, alors, d'après la page 167, $X'_n - X''_n$ converge aussi vers zéro en probabilité.

Enfin, si S′ et S″ sont presque certainement asymptotiques à S, les probabilités respectives que $|X'_n - X_n|$ tendent vers zéro et que $|X''_n - X|$ tende vers zéro sont égales à un. Si H et K sont ces deux événements, on a, d'après la page 28,

$$\Pr\{H\,K\} = \Pr.H. = 1.$$

Or, l'événement H. K, c'est-à-dire la convergence simultanée de $X'_n - X_n$ et de $X_n - X''_n$ vers zéro entraînant celle de $X'_n - X''_n$ vers zéro, celle-ci aura bien lieu presque certainement.

Supposons les variables aléatoires X_n indépendantes et soit $b_1, b_2, \ldots$ une suite de nombres certains. Les différences $X_n - b_n$ étant aussi

indépendantes. il en résulte (p. 247) que la probabilité de la convergence de la suite s'' des $X_n - b_n$ ou de la série $\Sigma(X_n - b_n)$ ne peut être égale qu'à 1 ou 0.

Considérons la suite s'' des $X_n - b_n$.

Dans le cas où la probabilité de sa convergence est égale à l'unité, on a vu (p. 249) que sa limite est un nombre certain b, par suite $X_n - b_n - b$ tend vers zéro presque certainement En posant $a_n = b_n + b$, on voit qu'alors, il existe une suite certaine $a_1, a_2, \ldots$ qui est asymptote à la suite $X_1, X_2, \ldots$.

Ainsi, pour une même suite s des X_n :

Ou bien quelle que soit la suite certaine des a_n la suite $X_n - a_n$ diverge presque certainement;

Ou bien il existe au moins une suite certaine $\sigma : a_1, a_2, \ldots$ qui est asymptote à la suite des X_n presque certainement. Et alors il existe une infinité de telles suites dont la plus générale s'obtient en ajoutant aux termes de σ les termes correspondants d'une suite quelconque de nombres certains convergeant vers zéro.

Notons enfin que les critères des pages 262 et 254 peuvent s'énoncer ainsi. Soient $X_1, X_2, \ldots$ des variables aléatoires indépendantes, $\overline{X_n}$ la valeur moyenne et μ_n l'écart quadratique moyen de X_n

1° Si la série certaine $\Sigma\mu_n^2$ est convergente, la série aléatoire ΣX_n est presque certainement asymptote à la série certaine $\Sigma\overline{X_n}$ de ses valeurs moyennes ($\Sigma\overline{X_n}$ pouvant être convergente ou divergente);

2° Si la série certaine $\Sigma\left(\frac{\mu_n}{n}\right)^2$ est convergente, la suite des moyennes arithmétiques aléatoires $v_n = \frac{X_1 + \ldots + X_n}{n}$ est presque certainement asymptote à la suite des moyennes arithmétiques certaines

$$V_n = \frac{\overline{X_1} + \ldots + \overline{X_n}}{n}.$$

PROBABILITÉS ÉGALES A 0 OU 1.

Théorème de Kolmogoroff. — Depuis le premier exemple, dû à M. Borel (p. 27), de probabilités égales nécessairement à zéro ou à l'unité, on a de plus en plus étendu le champ des événements qui présentent cette particularité : nous l'avons vu pages 246, 247.

Allant plus loin, M. Kolmogoroff (4, p. 60 à 61) a démontré le théorème suivant :

Soit $f(x_1, x_2, \ldots, x_n, \ldots)$ une fonction d'une infinité de variables numériques, qui est une fonction de Baire, c'est-à-dire qui est un polynome ou limite de polynomes ou limite de limites de polynomes, etc. Soit, maintenant, F l'événement consistant en ce que

$$f(X_1, X_2, \ldots, X_n, \ldots) = 0$$

$X_1, X_2, \ldots$, étant des variables aléatoires définies dans une suite d'épreuves (X_j déterminé à la $j^{\text{ième}}$ épreuve et la loi de probabilité de X_j pouvant dépendre des résultats des épreuves précédentes). Supposons que F ait une probabilité déterminée. Supposons aussi qu'il y ait une probabilité déterminée de F dans la catégorie des suites d'épreuves où $X_1, \ldots, X_n$ ont pris des valeurs déterminées $x_1, \ldots x_n$. Supposons enfin que ces deux probabilités qu'on peut désigner par Pr. F et $\text{Pr.}_{x_1, \ldots, x_n}\text{F}$ soient égales quel que soit n et quelles que soient les valeurs $x_1, \ldots, x_n$. *Sous ces hypothèses*, M. Kolmogoroff démontre que Pr. F *est nécessairement égal à zéro ou* 1.

Les hypotheses de ce théorème seront vérifiées d'elles-mêmes dans le cas particulier où les X_j sont indépendants et où de plus la fonction $f(x_1, \ldots x_n, \ldots)$ est une fonction de Baire ne changeant pas de valeur quand on modifie de façon arbitraire un nombre fini quelconque de ses variables $x_1, \ldots, x_n$.

On peut exprimer autrement le théorème de Kolmogoroff :

Considérons un événement E, défini dans la catégorie complexe C d'épreuves dont chacune consiste en une suite infinie des épreuves initialement considérées, l'événement E se traduisant par une certaine propriété **Π** de la suite x des valeurs $x_1, x_2, \ldots$ prises par $X_1, X_2, \ldots$. Appelons maintenant $\varphi(x_1, x_2, \ldots, x_n, \ldots)$ une fonction égale à zéro quand la suite x possède la propriété **Π** et à 1 dans le cas contraire.

Pour pouvoir appliquer la proposition à la fonction φ, il faut que la propriété **Π** soit telle que

$$\text{Pr. E} = \text{Pr.}\{\varphi = 0\} = \text{Pr.}_{x_1, \ldots, x_n}\{\varphi = 0\} = \text{Pr.}_{x_1, x_2, \ldots, x_n}\text{E},$$

c'est-à-dire que la probabilité de **Π** soit égale à la propriété que **Π** ait lieu quand $X_1 = x_1, \ldots, X_n = x_n$ et ceci quels que soient $n, x_1, \ldots, x_n$.

$\mathrm{Pr.}_{x_1, x_2, \ldots, x_n}\mathrm{E}$ *pour que* E *ait lieu quand* $X_1 = x_1, \ldots, X_n = x_n$, *alors, sauf peut-être si* x *réalise un évènement auxiliaire* e *de probabilité nulle sur* C, *on a, pour toute suite* x RÉALISANT E

$$\mathrm{Pr.\,E} = \lim_{n \to \infty} \mathrm{Pr.}_{x_1, \ldots, x_n}\mathrm{E}$$

quand $x_1, \ldots, x_n$ *appartient à* x.

La démonstration de ce théorème fait appel à plusieurs résultats intéressants en eux-mêmes et qui nécessitent quelques développements. Nous la renverrons au Second Livre.

SUPPLÉMENT MATHÉMATIQUE.

RAPPEL DES PROPRIÉTÉS DES FONCTIONS MONOTONES

Définition. — On appelle *fonctions monotones*, les fonctions non croissantes et les fonctions non décroissantes.

Si l'on fait tendre x vers une limite a, toujours dans le même sens et si $M(x)$ est monotone, $M(x) - M(a)$ gardera toujours le même signe sans s'éloigner de zéro et par suite aura une limite. On désigne par $M(a-0)$ et $M(a+0)$ les limites de $M(x)$ quand x tend vers a respectivement sans décroître et sans croître. Il est clair que $M(a)$ est compris entre $M(a-0)$ et $M(a+0)$. Il n'y aura discontinuité en a que si $M(a+0) - M(a-0) > 0$, cette quantité sera, dans ce cas, dite le *saut de discontinuité* de $M(x)$ au point a. Il est la somme du *saut à gauche* $M(a) - M(a-0)$ et du *saut à droite* $M(a+0) - M(a)$. On voit que les fonctions monotones ne possèdent pas les discontinuités les plus générales : leurs points de discontinuité sont tous « de première espèce ». En outre, l'ensemble des points de discontinuité d'une fonction monotone est « dénombrable », c'est-à-dire qu'on peut tous les numéroter au moyen d'entiers distincts. En effet, il est clair que la somme d'un nombre fini quelconque de ces sauts de discontinuité dans un intervalle fini (a, b) est, en valeur absolue, au plus égale à $|M(b) - M(a)|$. Il n'y a donc dans (a, b) qu'un nombre fini de points dont les sauts soient supérieurs à $\frac{1}{n}$. En les numérotant quand n passe de p à $p+1$, puis faisant croître p, on parviendra à numéroter tous ceux compris dans a, b. L'ensemble des points de discontinuité compris dans $(-n, +n)$ étant dénombrable pour tout entier n, il en est de même de tous les points de discontinuité.

Décomposition. — On peut décomposer toute fonction monotone, en deux ou trois parties de natures différentes, d'une façon qui fait mieux comprendre la structure de ces fonctions.

Une fonction monotone bornée $M(x)$, ayant un ensemble dénombrable de points de discontinuité $d_1, d_2, \ldots, d_n, \ldots$, posons

$$s'_i = s(d_i) - s(d_i - 0),$$
$$s''_i = s(d_i + 0) - s(d_i).$$

On appelle *fonction des sauts* de $M(x)$, la fonction

$$S(x) = \sum_{x_i \leq x} s'_i + \sum_{x_i < x} s''_i.$$

Comme les s'_i et s''_i sont tous de même signe et que la somme d'un nombre quelconque de ces quantités est au plus égale en valeur absolue à

$$|M(+\infty) - M(-\infty)|,$$

les deux séries du second membre sont absolument convergentes et définissent une fonction $S(x)$, monotone comme $M(x)$, et de même sens. En posant maintenant

$$C(x) = M(x) - S(x),$$

on voit que si $x' < x''$

$$C(x'') - C(x') = M(x'') - M(x') - \sum_{x' < x_i \leq x''} s'_i - \sum_{x' \leq x_i < x''} s''_i$$

$$= M(x'') - M(x') - [M(x'') - M(x'' - 0)]$$

$$- [M(x' + 0) - M(x')] - \sum_{x' < x_i < x''} s_i,$$

en posant $s_i = s'_i + s''_i$. Il est clair que $C(x)$ est monotone comme $M(x)$ et de même sens.

Quand x' étant fixe, x'' tend vers x' par valeurs supérieures, $C(x'')$ tend vers une limite $C(x' + 0)$. En prenant pour x'' seulement des points de continuité de $M(x)$, ce qui est évidemment possible, $M(x'') - M(x'' - 0)$ restera nul. Or $\sum\limits_{x' < x_i < x''} s_i$ est inférieur en valeur absolue à la valeur absolue du reste de la série absolument convergente Σs_i, pris à partir de la valeur n de i, telle que, à l'intérieur de $x'x''$, ne figurent ni d_1, ni d_2, .., ni d_n. Quand x'' tend vers x', n croît indéfiniment, donc $\sum\limits_{x' < x_i < x''}$ tend vers zéro. Dès lors, on voit que $C(x)$ est continue.

Dans le cas où $M(x)$ n'est pas borné, on peut prendre un nombre a arbitraire et poser

$$S(x) - S(a) = \sum_{a < x_i \leq x} s'_i + \sum_{a \leq x_i < x} s''_i \qquad \text{pour } x > a,$$

$$S(a) - S(x) = \sum_{x < x_i \leq a} s'_i + \sum_{x \leq x_i < a} s''_i \qquad \text{pour } x < a,$$

puis

$$C(x) = M(x) - M(a) - [S(x) - S(a)].$$

Alors le raisonnement précédent appliqué à un intervalle ab montrera que $C(x)$ est continu dans ab quel que soit b, c'est-à-dire partout.

En résumé, toute fonction monotone est la somme de deux fonctions monotones, l'une qui est sa fonction des sauts, l'autre qui est continue.

M. de La Vallée-Poussin a même démontré que cette dernière peut se décomposer à son tour en somme de deux fonctions monotones continues dont l'une est en outre une intégrale indéfinie et l'autre a une dérivée nulle « presque partout ».

M. Borel a donné dans le présent Traité, un exemple (I, I, p. 117) d'une fonction des probabilités totales qui est du second type c'est-à-dire qui est continue et telle que sa dérivée soit nulle sauf en un ensemble de points qu'on peut enfermer dans une suite d'intervalles dont la somme des longueurs aussi petite que l'on veut.

Limite d'une suite. — M. Montel a démontré pour les fonctions monotones continues une propriété importante qui a été ensuite étendue par M. Helly au cas des fonctions monotones quelconques, et qui nous a servi dans ce volume : *De toute suite infinie* S *de fonctions monotones, qui sont en chaque point, bornées dans leur ensemble, on peut extraire une suite infinie* σ *convergente.*

Il y a nécessairement dans S une suite s de fonctions monotones dans le même sens, soient $C_1(x)$, $C_2(x)$, ..., $C_n(x)$, ... Il suffit d'extraire σ de s. Supposons par exemple, les $C_n(x)$ non décroissantes.

Observons d'abord que si l'on considère une suite infinie de fonctions, monotones ou non, $f_p(x)$ qui sont bornées dans leur ensemble en chaque point x_n d'un ensemble dénombrable D de points x_1, x_2, ..., on peut extraire de cette suite de fonctions, une suite qui converge *sur* D. On peut en effet d'après un théorème connu de Weierstrass-Bolzano extraire de la suite des $f_n(x)$ une suite $f_1^{(1)}(x), f_2^{(1)}(x)$, ... qui converge au point x_1 et en général de la suite $f_1^{(r)}(x), f_2^{(r)}(x)$, ... une suite convergente au point x_r. La suite

$$f_1^{(1)}(x),\quad f_2^{(2)}(x),\quad \dots,\quad f_r^{(r)}(x),\quad \dots$$

est bien alors une suite extraite des $f_p(x)$ et qui converge en chaque point de D.

Appliquons cette observation à la suite s en prenant pour D l'ensemble dénombrable constitué des nombres rationnels. On pourra donc extraire de la suite s, une suite σ_0 de fonctions $c_i(x) \equiv C_{n_i}(x)$ qui converge sur D.

Soit ξ une abscisse appartenant ou non à D. Elle est comprise entre des abscisses rationnelles α, β. Lorsque i croît, les nombres $c_i(\alpha)$, $c_i(\beta)$ tendent vers deux limites qu'on peut appeler $c(\alpha)$, $c(\beta)$. Et si $\alpha < \beta$, alors $c_i(\alpha) \leqq c_i(\xi) \leqq c_i(\beta)$ et, par suite, $c(\alpha) \leqq c(\beta)$. Donc, toutes les limites de $c_i(\xi)$ quand i croît, sont aussi comprises entre $c(\alpha)$, $c(\beta)$. Or, $c(x)$ étant bien déterminé et non décroissant, sur D, la fonction $\varphi(\xi)$ égale à la borne supérieure des $c(\alpha)$ pour α *pris sur* D et $< \xi$, est bien déterminée pour tout ξ. Cette fonction est évidemment non décroissante; elle n'a donc qu'un ensemble

dénombrable de discontinuités. Si ξ n'est pas un point de discontinuité de $\varphi(x)$, on peut prendre $\eta < \xi < \zeta$ de sorte que $\varphi(\zeta) - \varphi(\eta)$ soit aussi petit que l'on veut. Si maintenant α et β sont deux points de D tels que $\eta < \alpha < \xi < \beta < \zeta$, on aura évidemment

$$\varphi(\eta) \leqq c(\alpha) \leqq c(\beta) \leqq \varphi(\zeta)$$

et comme toutes les limites de $c_1(\xi)$, $c_2(\xi)$, ... sont comprises entre $c(\alpha)$ et $c(\beta)$, elles le sont entre $\varphi(\eta)$ et $\varphi(\zeta)$. Elles sont par suite aussi voisines de $\varphi(\xi)$ que l'on veut. Ainsi la suite des $c_i(\xi)$ converge vers $\varphi(\xi)$ au moins en tout point de continuité de $\varphi(\xi)$. Ceci étant, on peut, comme on l'a vu, extraire de la suite σ_0 des c_i une suite σ des fonctions γ_1, γ_2, ... qui converge sur l'ensemble dénombrable des discontinuités de φ. Finalement, la suite σ, extraite de S est partout convergente.

On peut étendre le théorème au cas des fonctions à variations bornées [1] : si $V_1(x)$, $V_2(x)$ ... est une suite de fonctions qui dans tout intervalle fini, sont à variations bornées, si ces fonctions sont uniformément bornées en chaque point, si leurs variations totales sont uniformément bornées dans tout intervalle fini, alors on peut extraire de cette suite, une suite convergente, dont la limite est à variation bornée dans tout intervalle fini.

Mais cette généralisation ne nous est pas utile dans cet ouvrage.

Continuité uniforme. — On sait qu'une fonction définie et continue pour toute valeur de x et qui est par suite uniformément continue sur tout intervalle fini peut ne pas être uniformément continue sur l'ensemble de toute la droite illimitée. Et cela, même si cette fonction est monotone, comme le montre le cas de la fonction x^3. Cependant, *toute fonction continue monotone et bornée est uniformément continue*. En effet, si A et B sont ses deux bornes, on peut, étant donné $\varepsilon > 0$, déterminer deux points a, b tels que si f est non décroissante

$$A \leqq f(x) < A + \frac{\varepsilon}{2} \qquad \text{pour } x \leqq a,$$

$$B - \frac{\varepsilon}{2} < f(x) \leqq B \qquad \text{pour } x \geqq b.$$

Alors on peut déterminer ω tel que

$$|f(x') - f(x'')| < \frac{\varepsilon}{2} \qquad \text{pour } |x' - x''| < \omega \text{ et } a \leqq {x' \atop x''} \leqq b$$

On voit alors qu'on aura, en prenant $\omega < b - a$,

$$|f(x') - f(x'')| < \varepsilon \qquad \text{pour } |x' - x''| < \omega,$$

quels que soient x', x'' *sur la droite illimitée*.

---

(1) On dit qu'une fonction $V(x)$ est à *variation bornée* sur un intervalle a, b si la somme $\sum_k |V(x_k) - V(x_{k-1})|$ garde une borne supérieure finie quand on fait varier le mode de division $a < x_1 < x_2 \ldots < x_n < b$ de a, b. Cette borne supérieure est la « variation totale » de $V(x)$ sur ab.

On peut se demander si la convergence d'une suite de fonctions monotones $\gamma_n(x)$ est une convergence uniforme. Il est facile de montrer que si la fonction limite $\gamma(x)$ est continue sur un segment (a, b), la convergence est uniforme sur ce segment. On peut en effet diviser un tel segment a, b en un nombre fini de segments x_{k-1}, x_k, sur chacun desquels l'accroissement de $\gamma(x)$ est inférieur à un nombre ε positif donné d'avance. Supposons, par exemple, non décroissantes, les fonctions $\gamma_n(x)$ et $\gamma(x)$. Pour n assez grand ($n > N$), on aura $|\gamma_n(x_k) - \gamma(x_k)| < \varepsilon$, à la fois pour les différentes valeurs de k. Si x appartient à (a, b), il appartient à l'un des intervalles (x_{k-1}, x_k), en posant $a = x_0$, $b = x_n$, et l'on a, pour $n > N$,

$$\gamma_n(x_{k-1}) - \gamma(x_k) \leq \gamma_n(x) - \gamma(x) \leq \gamma_n(x_k) - \gamma(x_{k-1}),$$

d'où

$$-2\varepsilon \leq [\gamma_n(x_{k-1}) - \gamma(x_{k-1})] + [\gamma(x_{k-1}) - \gamma(x_k)]$$
$$\leq \gamma_n(x) - \gamma(x) \leq [\gamma_n(x_k) - \gamma(x_k)] + [\gamma(x_k) - \gamma(x_{k-1})] \leq 2\varepsilon.$$

Si $\gamma(x)$ est continu sur toute la droite, la convergence est uniforme sur tout intervalle fini. Il n'en résulte pas que la convergence soit uniforme sur toute la droite, même si l'on suppose bornée la fonction limite, comme le montre le cas où $\gamma_n(x) = \frac{x^3}{n}$. Si la fonction-limite $\gamma(x)$ est continue et bornée, la convergence ne pourra être uniforme que si les fonctions $\gamma_n(x)$ de la suite sont elles-mêmes bornées à partir d'un certain rang et si les bornes de $\gamma_n(x)$ tendent respectivement vers celles de $\gamma(x)$. On le voit par exemple, en prenant comme nouvelle variable $y = \operatorname{th} x$ de façon à ramener au cas d'un intervalle fini $-1 < y < +1$. Cette transformation montre en même temps que les conditions indiquées sont suffisantes.

Généralisation. — Abandonnons l'hypothèse que $\gamma(x)$ soit une fonction continue pour l'hypothèse plus générale que cette fonction soit bornée. Dire que les fonctions $\gamma_n(x)$ convergent uniformément vers $\gamma(x)$ sur tout l'intervalle de variation de x (soit de $-\infty$ à $+\infty$), c'est dire qu'à tout ε positif correspond un entier N tel que l'on ait, quel que soit x_0,

$$(1) \qquad |\gamma_n(x_0) - \gamma(x_0)| < \varepsilon \quad \text{pour} \quad n > N.$$

1° Il en résulte d'abord que les fonctions $\gamma_n(x)$ sont bornées, à partir d'un certain rang, et puisque ce sont des fonctions monotones, que $\gamma_n(-\infty)$ et $\gamma_n(+\infty)$ ont des valeurs finies. De plus, en vertu de l'inégalité ci-dessus, on aura

$$|\gamma_n(-\infty) - \gamma(-\infty)| < \varepsilon \quad \text{et} \quad |\gamma_n(+\infty) - \gamma(+\infty)| < \varepsilon \quad \text{pour} \quad n > N.$$

Donc, $\gamma_n(\pm\infty)$ tend vers $\gamma(\pm\infty)$.

2° En faisant tendre x_0 vers x par valeurs plus grandes ou par valeurs

plus petites dans (1), on aura

$$|\gamma_n(x-0)-\gamma(x-0)|<\varepsilon, \quad |\gamma_n(x+0)-\gamma(x+0)|<\varepsilon \quad \text{pour} \quad n>\mathrm{N}.$$

Donc, pour chaque valeur de x, non seulement $\gamma_n(x)-\gamma(x)$, mais aussi $\gamma_n(x\pm 0)-\gamma(x\pm 0)$ tendent vers zéro.

Dans le cas où les $\gamma_n(x)$ et $\gamma(x)$ sont des fonctions des probabilités totales, M. Cantelli (9, p. 7), a démontré un théorème qui, lorsqu'on explicite les hypothèses de son énoncé, est équivalent à la réciproque de la condition nécessaire 2° qui vient d'être formulée (la condition 1° étant alors remplie d'elle-même). Sa démonstration s'étend sans peine au cas où les fonctions $\gamma_n(x)$ restant non décroissantes, on suppose seulement $\gamma(x)$ bornée. Supposons alors les conditions 1° et 2° vérifiées.

Soit ω un nombre positif arbitraire. On peut déterminer un nombre fini de points $a_1 < a_2 < \ldots < a_r$ tels qu'*à l'intérieur* des segments finis ou infinis déterminés par ces points, $\gamma(x)$ varie de moins de ω. En effet, on peut d'abord déterminer a_1 et a_r, de sorte que

$$(2) \qquad \gamma(a_1)-\gamma(-\infty)<\omega, \qquad \gamma(+\infty)-\gamma(a_r)<\omega.$$

Décomposons maintenant $\gamma(x)$ en la somme $s(x)+\varphi(x)$ de sa fonction des sauts $s(x)$ et d'une fonction continue $\varphi(x)$.

On peut intercaler des points de division dans (a_1, a_r), de sorte que $\varphi(x)$ varie dans chacun de moins de $\frac{\omega}{2}$. On formera la suite annoncée en adjoignant aux points de division ainsi fixés des points choisis de telle façon que $s(x)$ varie de moins de $\frac{\omega}{2}$ à l'*intérieur* des segments qu'ils déterminent. Il suffit de prendre pour ces derniers points ceux des points de discontinuités $d_1, d_2, \ldots$ de $\gamma(x)$, qui, d'une part, sont compris entre a_1 et a_r, d'autre part, sont de rangs $\leqq \mathrm{N}$, N étant choisi de telle façon que le reste de rang N : $s_{\mathrm{N}+1}+s_{\mathrm{N}+2}+\ldots$ de la série convergente Σs_n des sauts de $\gamma(x)$, soit $<\frac{\omega}{2}$.

On a donc maintenant, en outre des inégalités (2),

$$\gamma(x'')-\gamma(x')<\omega \qquad \text{pour} \qquad a_j<x'\leqq x''<a_{j+1}$$

et, par suite,

$$\gamma(a_{j+1}-0)-\gamma(a_j+0)\leqq\omega.$$

Ceci étant, prenons q assez grand pour que, pour $n>q$, toutes les quantités $|\gamma_n(-\infty)-\gamma(-\infty)|$, $|\gamma_n(a_j\pm 0)-\gamma(a_j\pm 0)|$, $|\gamma_n(a_j)-\gamma(a_j)|$, $|\gamma_n(+\infty)-\gamma(+\infty)|$ soient $<\omega$. Pour tout x intérieur à l'un des intervalles finis ou infinis considérés, on aura $|\gamma_n(x)-\gamma(x)|<2\omega$ pour $n>q$, car, si, par exemple, $a_j<x<a_{j+1}$, on aura

$$\begin{aligned}-2\omega<[\gamma_n(a_j+0)-\gamma(a_j+0)]+[\gamma(a_j+0)-\gamma(a_{j+1}-0)]&\leqq\gamma_n(x)-\gamma(x)\\ \leqq[\gamma_n(a_{j+1}-0)-\gamma(a_{j+1}-0)]+[\gamma(a_{j+1}-0)-\gamma(a_j+0)]&<2\omega.\end{aligned}$$

Dès lors, on aura $|\gamma_n(x)-\gamma(x)| < 2\omega$, quel que soit x, quand n est supérieur à un nombre q indépendant de x : $\gamma_n(x)$ converge uniformément vers $\gamma(x)$.

Nous avons ainsi obtenu le résultat suivant, qui étend un peu et complète le théorème de Cantelli cité ci-dessus, lequel étendait lui-même un théorème de Polya.

Soient $\gamma_n(x)$ des fonctions non décroissantes. Pour qu'elles convergent uniformément sur l'intervalle $(-\infty, +\infty)$ vers une fonction bornée $\gamma(x)$, il faut et il suffit : que chacune des fonctions $\gamma_n(x)$, $\gamma_n(x+0)$, $\gamma_n(x-0)$ converge pour chaque valeur de x vers la fonction correspondante $\gamma(x)$, $\gamma(x+0)$, $\gamma(x-0)$, que les fonctions $\gamma_n(x)$ soient bornées à partir d'un certain rang n et que les bornes $\gamma_n(+\infty)$, $\gamma_n(-\infty)$ convergent vers les bornes correspondantes $\gamma(+\infty)$ et $\gamma(-\infty)$.

Les conditions seraient moins strictes s'il s'agissait seulement de réaliser la convergence uniforme sur tout segment fini. On voit encore, comme plus haut, que les fonctions $\gamma_n(x)$, $\gamma_n(x \pm 0)$ devraient converger respectivement vers $\gamma(x)$ et $\gamma(x \pm 0)$, quel que soit x. Même, en abandonnant l'hypothèse que $\gamma(x)$ est borné, ces seules conditions sont suffisantes. Pour montrer que la convergence est uniforme sur un segment fini arbitraire donné (α, β), il suffit, en effet, d'appliquer le théorème ci-dessus aux fonctions $\Lambda_n(x)$, $\Lambda(x)$ coïncidant avec $\gamma_n(x)$ et $\gamma(x)$ pour $\alpha \leqq x \leqq \beta$ et continues et constantes en dehors de ce segment.

NOTE A

UNE PROPRIÉTÉ NOUVELLE DE LA SECONDE LOI DE LAPLACE.

Une hypothèse de M. Paul Lévy. — Soit Z la somme de deux variables aléatoires indépendantes X, Y. On sait (ce Traité, I, 2, p. 51) que si X et Y obéissent à la seconde loi de Laplace (la loi faisant intervenir deux paramètres qui peuvent être respectivement différents pour X et pour Y), il en est de même de Z. Il est clair que la conclusion subsiste si l'une des variables X, Y obéit à la seconde loi de Laplace, l'autre restant certaine (sauf peut-être dans le cas où se produit un événement de probabilité nulle).

M. Paul Lévy (5) a émis l'hypothèse que la réciproque est exacte *si la somme Z de deux variables aléatoires indépendantes* X, Y *obéit à la seconde loi de Laplace, il en est de même de* X *et de* Y. Bien entendu, ceci suppose écarté le cas exceptionnel où l'une des variables X, Y obéissant encore à la seconde loi de Laplace, l'autre resterait certaine (ou « presque certaine »).

La démonstration de M. Cramér. — L'exactitude de cette proposition vient d'être démontrée par M. H. Cramér (3). Nous allons reproduire, à quelques légeres modifications pres, sa démonstration basée sur l'emploi des fonctions caractéristiques. (Il nous paraît d'ailleurs probable qu'on pourrait aussi arriver au même résultat par la considération des moments, ce qui serait souhaitable pour un enseignement élémentaire où l'on ne pourrait faire appel, comme M. Cramér, à la théorie des fonctions analytiques entieres.)

Nous décomposerons la démonstration de M. Cramér en deux parties. Une première partie, conçue en vue du problème actuel, peut être cependant présentée sous la forme d'un lemme indépendant

de ce problème. Elle conduit à formuler un ensemble de deux conditions, assez hétéroclites il est vrai, mais utiles ici, ensemble qui est à la fois nécessaire et suffisant pour qu'une variable aléatoire absolument quelconque X obéisse à la seconde loi de Laplace. La seconde partie applique ce résultat à l'hypothèse de M. Lévy.

Lemme. — Nous ferons une remarque préliminaire concernant la fonction caractéristique $\varphi(z) = \mathfrak{M} e^{izX}$ d'une variable aléatoire X, variable quelconque, pour le moment sans rapport avec le problème actuel. Observons d'abord que si U, V sont deux variables aléatoires réelles, en posant $W = U + iV$ (et par définition)

$$\mathfrak{M} W = \mathfrak{M} U + i \mathfrak{M} V,$$

on a

$$|\mathfrak{M} W|^2 = (\mathfrak{M} U)^2 + (\mathfrak{M} V)^2 \leqq \mathfrak{M} U^2 + \mathfrak{M} V^2 = \mathfrak{M}(U^2 + V^2),$$

d'où

$$|\mathfrak{M} W|^2 \leqq \mathfrak{M} |W|^2.$$

Donc, pour z réel ou complexe,

$$|\varphi(z)|^2 \leqq \mathfrak{M} |e^{izX}|^2 = \mathfrak{M} |e^{2izX}|.$$

Or, pour tout nombre réel ou complexe t, on a $|e^t| \leqq e^{|t|}$, d'autre part, on a évidemment $2|zX| \leqq |z|^2 + X^2$, d'où

$$|\varphi(z)|^2 \leqq \mathfrak{M} e^{2|zX|} \leqq \mathfrak{M} e^{|z|^2 + X^2} = e^{|z|^2} \mathfrak{M} e^{X^2}.$$

D'ailleurs, comme on peut écrire $izX = i\frac{z}{c} cX$, quel que soit le nombre réel certain $c \neq 0$, on aurait de la même manière

$$(1) \qquad |\varphi(z)|^2 \leqq e^{\left|\frac{z}{c}\right|^2} \mathfrak{M} e^{c^2X^2}.$$

Une telle inégalité n'aura d'ailleurs d'utilité que si X est tel que, pour au moins une valeur a de c, la valeur moyenne de $e^{c^2X^2}$ est finie. *Faisons cette hypothèse.*

Dans ce cas, comme

$$(2) \qquad \frac{(a^2X^2)^n}{n!} \leqq e^{a^2X^2},$$

quel que soit l'entier n, les moments

$$\mathfrak{M} \frac{a^{2n}X^{2n}}{n!} = \frac{a^{2n}}{n!} \mathfrak{M} X^{2n}$$

seront tous finis, et puisque (p. 60) $\sqrt[p]{\mathfrak{M}|X|^p}$ est une fonction non décroissante de p, $\mathfrak{M}|X|^p$ et, par suite, $\mathfrak{M}X^p$, seront finis quel que soit p. On pourra donc attacher une signification aux termes du développement formel

$$(3)\qquad \varphi(z) = 1 + \sum_{p=1}^{p=+\infty} \frac{(iz)^p}{p!}\,\mathfrak{M}X^p.$$

Pour étudier la convergence de cette série, on peut trouver une série majorante de $\varphi(z)$. On tire en effet de (2), en posant $K = \mathfrak{M}e^{a^2X^2}$,

$$\mathfrak{M}X^{2n} \leqq \frac{n!}{a^{2n}}\,K.$$

En appelant $\varphi_1(z)$ et $\varphi_2(z)$ les sommes des termes de degrés respectivement pairs et impairs de $\varphi(z)$, on a donc

$$(4)\qquad \varphi_1(z) \ll 1 + \sum_{n=1}^{n=+\infty} \frac{|z|^{2n}}{(2n)!}\,\frac{n!}{a^{2n}}\,K = 1 + \sum_{n=1}^{n=+\infty} \left|\frac{z}{a}\right|^{2n} K\,\frac{n!}{(2n)!}$$

et puisque

$$|\mathfrak{M}X^{2n-1}| \leqq \mathfrak{M}|X|^{2n-1} \leqq [\mathfrak{M}X^{2n}]^{\frac{2n-1}{2n}}$$

on aura

$$(5)\qquad \varphi_2(z) \ll \sum_{n=1}^{n=+\infty} \frac{|z|^{2n-1}}{(2n-1)!}\left[\frac{n!}{a^{2n}}\,K\right]^{\frac{2n-1}{2n}}$$

(Ici le signe $\ll$ indique, comme d'habitude, que les termes de la série à droite majorent en modules ceux de la série à gauche.)

Or, on voit facilement que les deux séries majorantes sont convergentes quel que soit z. Ainsi $\varphi(z)$ est une fonction *entière* de z. De plus, cette fonction entière vérifie, quel que soit z, la relation (1) pour $c = a$; d'où

$$(6)\qquad |\varphi(z)| \leqq e^{\frac{|z|^2}{2a^2}}\sqrt{K}.$$

Ce résultat nous invite à considérer un cas particulier, où l'on pourra obtenir la forme explicite de $\varphi(z)$. C'est celui où $\varphi(z)$ n'a pas de racine. Si une fonction entière n'a pas de racine, elle peut [1] se mettre sous la forme $\varphi(z) = e^{G(z)}$ où $G(z)$ est une fonction entière. On peut alors opérer de deux manières :

---

[1] *Voir* par exemple GOURSAT, *Cours d'Analyse mathématique*, t. II, p. 156.

1° Ou bien, comme M. Cramér, recourir à la notion d'ordre [1] d'une fonction entière. Alors, en vertu de (6), $\varphi(z)$ est d'ordre $\leqq 2$ et $\varphi(z)$ n'ayant pas de racine, $\varphi(z) = e^{G(z)}$ où $G(z)$ devra être un polynome de degré 2 au plus ;

2° Ou bien, pour éviter d'avoir recours dans cet Ouvrage à une théorie qui dépasse les éléments de la théorie des fonctions analytiques, on peut raisonner directement. Si l'on pose

$$G(z) = G(x+iy) = P(x, y) + iQ(x, y),$$

on aura

$$|\varphi| = e^{P}$$

et, par suite.

$$P(x, y) \leqq \frac{x^2+y^2}{2a^2} + k \qquad \text{avec} \quad k = \mathcal{L}\sqrt{K}$$

Donc la fonction harmonique $U(x, y) \equiv P(x, y) - k$ est partout majorée par $\dfrac{x^2+y^2}{2a^2}$.

Or, on sait que U est développable en une série (uniformément convergente dans tout domaine fini) de polynomes homogènes d'ordre croissant

$$U(x, y) = U_0(x, y) + \ldots + U_n(x, y) + \ldots$$

Il est alors à peu près évident que U doit se réduire au second degré, sans quoi $|U|$ croîtrait plus vite que toute puissance m de $r = \sqrt{x^2+y^2}$ (s'il y a dans les U_n un terme U_n non $\equiv 0$ de degré $n > m$), et en particulier plus vite que $\dfrac{r^2}{2a^2}$.

On peut le prouver ainsi : en coordonnées polaires, r, θ, le développement en série de Fourier de U pour r constant, étant la partie réelle d'une série entière en $z = re^{i\theta}$, est de la forme

$$U = U_0 + r(a_1 \cos\theta + b_1 \sin\theta) + \ldots + r^n(a_n \cos n\theta + b_n \sin n\theta) + \ldots,$$

où les a_n, b_n sont des constantes, le terme général n'étant autre que $U_n(x, y)$. On a donc, suivant une remarque de M. Cramér,

$$2\pi U_0 \pm \pi r^n a_n = \int_0^{2\pi} U(1 \pm \cos n\theta)\, d\theta \leqq \frac{r^2}{a^2} 2\pi ;$$

[1] *Voir*, par exemple, VALIRON, *Fonctions entières et fonctions méromorphes* (*Mémorial des Sc. math.*, Gauthier-Villars, Paris, 1925, p. 24, théorème XXIX).

d'où

$$\frac{2U_0}{r^n} \pm a_n \leqq \frac{2}{a^2 r^{n-2}}.$$

En faisant croître r, on voit que, pour $n > 2$, on obtient à la limite $\pm a_n \leqq 0$ D'où $a_n = 0$, et, de même, $b_n = 0$ pour $n > 2$, d'où

$$U_n(x, y) \equiv 0 \qquad \text{pour} \quad n > 2.$$

Ainsi U, et par suite $P(x, y)$, sont au plus du second degre

Comme d'ailleurs, on a

$$\frac{\partial Q}{\partial y} = \frac{\partial P}{\partial x}, \quad \frac{\partial Q}{\partial x} = -\frac{\partial P}{\partial y},$$

Q est nécessairement aussi du second degré en x, y et par suite $G(z)$ est du second degré en z.

Alors $\varphi(z)$ sera de la forme

$$(7) \qquad \varphi(z) = e^{\alpha z^2 + \beta z + \gamma},$$

où α, β, γ sont des constantes. Et puisque $\varphi(0) = 1$, on aura $\gamma = 0$.

D'ailleurs, il résulte du développement (3) qu'on a

$$\varphi^{(p)}(0) = i^p \mathfrak{M} X^p$$

D'où, en raison de (7),

$$\beta = i\mathfrak{M}X, \quad \beta^2 + 2\alpha = i^2 \mathfrak{M} X^2.$$

Donc

$$\beta = i\overline{X}, \qquad 2\alpha = \overline{X}^2 - \mathfrak{M}X^2 = -\mathfrak{M}(X - \overline{X})^2 = -\mu^2,$$

μ étant l'écart quadratique moyen de X. Des lors,

$$\varphi(z) = e^{-\frac{\mu^2}{2}z^2 + i\overline{X}z}.$$

On sait, (I, 2, p. 29 et 45) qu'alors, si $\mu \neq 0$, la fonction des probabilités totales correspondante est

$$F(x) = \frac{1}{\mu\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\overline{X})}{2\mu^2}} dt;$$

c'est la seconde loi de Laplace.

Ainsi, pour qu'une variable aléatoire X qui n'est pas « presque certaine » obéisse à cette loi, il suffit que la valeur moyenne de $e^{c^2X^2}$

soit finie pour au moins une valeur a de la constante réelle c et que la fonction caractéristique de X n'ait pas de racines.

Ces conditions suffisantes sont aussi nécessaires. Car, si X vérifie la seconde loi de Laplace : 1° sa fonction caractéristique est de la forme

$$\varphi(z) = e^{-\frac{\mu^2}{2}z^2 + iz\bar{X}}.$$

$\varphi(z)$ n'a donc pas de racines;

2° La densité de probabilité est de la forme

$$f(x) = \frac{1}{\mu\sqrt{2\pi}} e^{-\frac{(x-\bar{X})^2}{2\mu^2}} dx,$$

on a donc

$$\mathcal{M} e^{c^2X^2} = \frac{1}{\mu\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\bar{X})^2}{2\mu^2} + c^2x^2} dx$$

L'exposant de e peut s'écrire, en prenant $c^2 < \frac{1}{2\mu^2}$ et posant

$$h^2 = \frac{1}{2\mu^2} - c^2,$$

$$-h^2x^2 + \frac{2x\bar{X}}{2\mu^2} - \frac{\bar{X}^2}{2\mu^2} = -h^2\left(x - \frac{\bar{X}}{2h^2\mu^2}\right)^2 - \frac{\bar{X}^2}{2\mu^2}\left[1 - \frac{1}{2h^2\mu^2}\right]$$

En posant $x - \frac{\bar{X}}{2h^2\mu^2} = y$, on aura

$$\mathcal{M} e^{c^2X^2} = R \int_{-\infty}^{+\infty} e^{-h^2y^2} dy,$$

où R est une constante finie, et où l'intégrale du second membre est convergente. Ainsi on voit que $\mathcal{M} e^{c^2X^2}$ est bien finie pour $c^2 < \frac{1}{2\mu^2}$.

Application du lemme à l'hypothèse de Lévy. — Il suffit maintenant de démontrer que si $Z = X + Y$ obéit à la seconde loi de Laplace et si X et Y sont indépendants. X vérifie nécessairement les deux conditions ci-dessus. Il en sera naturellement de même de Y.

Nous observerons d'abord en modifiant sur ce point, avec M. Paul Lévy, la démonstration de M. Cramér, qu'on peut établir une relation

entre les fonctions des probabilités totales $F(x)$, $G(x)$, $H(x)$ de X, Y, $Z = X + Y$, quand on prend pour origine une médiane de Y. Alors on a (p. 40)

$$G(o) \leqq \frac{1}{2} \leqq G(+o)$$

Or, X et Y étant indépendants, le produit $F(x+o)\,G(+o)$ est égal à la probabilité, qu'on ait à la fois $X \leqq x$, $Y \leqq o$, probabilité évidemment au plus égale à la probabilité que $X + Y \leqq x$, c'est-à-dire à $H(x+o)$. On a donc

$$H(x+o) \geqq F(x+o)\,G(+o) \geqq \frac{1}{2} F(x+o)$$

Pour x_0 donné, prenons $x < x_0$, on aura

$$F(x) \leqq F(x+o) \leqq 2H(x+o) \leqq 2H(x_0)$$

D'où, en faisant croître x vers x_0,

$$F(x_0) = F(x_0 - o) = \lim_{x \to x_0} F(x) \leqq 2H(x_0)$$

Ainsi on a toujours $F(x) \leqq 2H(x)$ quel que soit x. On verrait de même que

$$1 - F(x) \leqq 2[1 - H(x)]$$

Or il s'agit d'abord de voir si l'intégrale

$$\mathfrak{M} e^{c^2 X^2} = \int_{-\infty}^{+\infty} e^{c^2 x^2}\, dF(x) = \int_0^{+\infty} + \int_{-\infty}^0 \tag{9}$$

est convergente. Commençons par la dernière intégrale. C'est la limite de

$$\int_{-A}^0 e^{c^2 x^2}\, dF(x) = [F(x)\,e^{c^2 x^2}]_{-A}^0 - \int_{-A}^0 F(x)\, de^{c^2 x^2} \tag{10}$$

On a

$$[F(x)\,e^{c^2 x^2}]_{-A}^0 = F(o) - F(-A)\,e^{c^2 A^2} = F(o) - 2\theta(A)\,H(-A)\,e^{c^2 A^2}$$

avec $o \leqq \theta(A) \leqq 1$; et

$$\int_{-A}^0 F(x)\, de^{c^2 x^2} = 2\theta_1(A) \int_{-A}^0 H(x)\, de^{c^2 x^2}$$
$$= 2\theta_1 [H(x)\,e^{c^2 x^2}]_{-A}^0 - 2\theta_1 \int_{-A}^0 e^{c^2 x^2}\, dH(x)$$

avec $o \leqq \theta_1(A) \leqq 1$.

Nous avons à montrer que le premier membre de (10) (qui ne peut décroître quand A croît) ne tend pas vers l'infini. Il suffit donc maintenant de prouver que

$$H(-A)\,e^{c^2A^2} \qquad \text{et} \qquad \int_{-A}^{0} e^{c^2x^2}\,dH(x)$$

restent bornés quand $A \to +\infty$, pourvu que c soit choisi convenablement. Or, on voit facilement qu'il en est ainsi quand Z obéit à la loi de Gauss.

En effet, la seconde quantité est au plus égale à $\mathfrak{M}\,e^{c^2Z^2}$ qui, d'après un raisonnement fait plus haut, est finie quand $c^2 < \frac{1}{2\rho^2}$ où ρ est l'écart quadratique moyen de Z. D'autre part,

$$H(x) = \frac{1}{\rho\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\bar{Z})^2}{2\rho^2}}\,dt.$$

En posant

$$\frac{x-\bar{Z}}{\rho\sqrt{2}} = u, \quad \frac{-A-\bar{Z}}{\rho\sqrt{2}} = -B, \quad \frac{t-\bar{Z}}{\rho\sqrt{2}} = s,$$

on a

$$H(\bar{Z} + u\rho\sqrt{2}) = \int_{-\infty}^{u} e^{-u^2}\,du = \Psi(u),$$

$$(11) \qquad H(-A)\,e^{c^2A^2} = \Psi(-B)\,e^{c^2(\bar{Z}-B\rho\sqrt{2})^2}$$

$$\Psi(-B) = \frac{1}{\sqrt{\pi}} \int_{B}^{+\infty} e^{-u^2}\,du = K(B),$$

et on a vu, p. 219, que

$$K(B) = \frac{1}{2B\sqrt{\pi}}\,e^{-B^2}\left[1 - \frac{\omega(B)}{2B^2}\right]$$

avec $0 < \omega(B) < 1$.

Le second membre de (11) est donc :

$$\frac{1}{2B\sqrt{\pi}}\left[1 - \frac{\omega(B)}{2B^2}\right] e^{c^2(\bar{Z}-B\rho\sqrt{2})^2 - B^2},$$

et, sous la même condition $c^2 < \frac{1}{2\rho^2}$, il reste borné et même tend vers zéro quand B croît.

En résumé, l'intégrale $\int_{-\infty}^{0}$ dans (9) est convergente pour

$$c^2 < \frac{1}{2\rho^2}.$$

On verrait de même que, pour les mêmes valeurs de c, l'intégrale $\int_0$ est convergente, et par suite que $\mathfrak{M}e^{2\lambda^2}$ est finie.

Il en résulte, en particulier, que $\varphi(z)$ est une fonction entière; et de même $\psi(z)$, en désignant par $\psi(z)$ et $\chi(z)$ les fonctions caractéristiques de Y et de $Z = X + Y$. Or, on a,

$$\varphi(z)\psi(z) = \chi(z),$$

d'autre part, Z obéissant par hypothèse à la seconde loi de Laplace, $\chi(z)$ n'a pas de racine. Mais, pour toute valeur finie de z, les fonctions entières $\varphi(z)$ et $\psi(z)$ sont finies; leur produit ne s'annulant pas, elles n'ont donc pas de racine.

Ainsi, l'exactitude de l'hypothèse de Lévy se trouve établie. Après l'avoir formulée, l'auteur en avait déduit plusieurs conséquences intéressantes pour lesquelles nous renvoyons à son mémoire (Paul Lévy, 5, Chap. III, IV). D'autre part, M. Cramér (3) a étendu la proposition au cas où X, Y sont deux points aléatoires (p. 82) de l'espace à ν dimensions.

NOTE B

DISTANCE DE DEUX VARIABLES ALÉATOIRES ET DISTANCE DE DEUX LOIS DE PROBABILITÉ

Par M. PAUL LÉVY.

Remarques préliminaires. — Si l'on connaît les lois de probabilité dont dépendent deux variables X et Y, on n'est pas pour cela à même de déterminer leur distance. Il faut, en outre, connaître la *corrélation* qui existe entre elles, c'est-à-dire la manière dont la connaissance de l'une influe sur la loi de probabilité de l'autre. La notion de corrélation n'ayant pas été introduite dans ce volume, montrons son importance par un exemple. Supposons que chacune des variables X et Y dépende de la loi de Laplace réduite. Cela est possible de bien des manières, il peut arriver que l'on ait sûrement $X = Y$, ou sûrement $X = -Y$, ou encore que X et Y soient des variables indépendantes. Une infinité d'autres hypothèses sont possibles. Pour celles que nous venons d'indiquer la différence $X - Y$ est nulle dans le premier cas, égale à $2Z$ dans le second et à $\sqrt{2}Z$ dans le troisième, Z étant une variable qui dépend, comme X et Y, de la loi de Laplace réduite.

Or, la distance des variables X et Y dépend essentiellement de la nature de la variable aléatoire $X - Y$, et ne dépend que d'elle, du moins pour les définitions simples de la distance indiquées par M. Fréchet (il semble que l'on puisse ajouter : pour toutes celles qui se présentent naturellement à l'esprit). On voit par l'exemple précédent combien cette distance peut varier avec la corrélation qui existe entre X et Y. Cette remarque est tout à fait générale; ce n'est que dans le cas où l'une des variables est (sauf dans des cas de probabilité

nulle) égale à un nombre déterminé, c'est-à-dire lorsqu'elle cesse d'être aléatoire, que la notion de corrélation disparaît et que la distance (X, Y) est déterminée par les lois dont dépendent ces variables

On voit donc que : si X est une variable véritablement aléatoire (non presque sûrement égale à une constante), pour que la variable Y soit très voisine de X, il est nécessaire, mais non suffisant, que la loi de probabilité de Y diffère très peu de celle de X; cette condition étant vérifiée, une condition supplémentaire, faisant intervenir la corrélation de X et Y, sera nécessaire et suffisante

Ces remarques nous conduisent à nous demander s'il est possible de définir la distance de deux lois L et L', et nous suggèrent même une définition possible : *cette distance* (L, L') *est la borne inférieure de la distance* (X, Y), X *et* Y *étant deux variables aléatoires dépendant respectivement des lois* L *et* L', *et dont la corrélation varie de toutes les manières possibles.*

Montrons que, *quelle que soit la définition adoptée pour la distance* (X, Y), *cette définition de* (L, L') *est acceptable*, c'est-à-dire que l'inégalité triangulaire

$$(X, Y) \leqq (X, Z) + (Y, Z) \tag{1}$$

entraîne l'inégalité analogue

$$(L, L') \leqq (L, L'') + (L', L'') \tag{2}$$

On peut, en effet, X, Y et Z dépendant respectivement des lois L, L', L'' et ε étant arbitrairement petit, établir *simultanément* entre X et Z et entre Y et Z des corrélations telles que

$$(X, Z) \leqq (L, L'') + \varepsilon, \qquad (Y, Z) \leqq (L', L'') + \varepsilon,$$

et par suite, d'après (1),

$$(X, Y) \leqq (L, L'') + (L', L'') + 2\varepsilon.$$

L'inégalité (2) en résulte.

Bien entendu, il faut aussi vérifier que, L étant fixe et L' variable, (L, L') tend vers zéro si L' tend vers L et dans ce cas seulement; cette vérification ne présente aucune difficulté [1].

---

[1] Si les lois L et L' varient toutes les deux, la question de savoir quand on doit

Ajoutons que la borne inférieure qui nous sert à définir (L, L′) est toujours effectivement atteinte (ce qui permet de supprimer ε dans la démonstration précédente). Sans démontrer ce point, indiquons qu'il résulte aisément du fait que les lois de probabilité à deux variables X, Y que l'on peut considérer lorsque les lois L et L′ sont connues, forment un ensemble compact.

Définitions directes de la distance de deux lois. — La définition qui précède présente l'inconvénient qu'il peut être très difficile de déterminer la distance de deux lois données par leurs fonctions des probabilités totales. Il est donc utile d'avoir des définitions directes. Nous allons en indiquer deux.

1° La première est celle que j'ai indiquée ou plutôt utilisée (sans prononcer le mot de distance) en 1925 (*Calcul des probabilités*, p. 199-200). Définissons la loi L dont dépend une variable aléatoire X par la courbe Γ représentant la fonction des probabilités totales

$$y = \Pr[X < x],$$

avec toutefois cette convention qu'à chaque valeur de x pour laquelle cette fonction est discontinue, nous ferons correspondre le segment

$$\Pr.[X < x] \leqq y \leqq \Pr[X \leqq x].$$

La courbe est alors continue, et manifestement coupée en un point et un seul par n'importe quelle parallèle à la droite

$$x + y = 0.$$

Deux lois L et L′ étant ainsi représentées par deux courbes Γ et Γ′ que la droite $x + y = c$ coupe respectivement en A et A′, la distance (L, L′) = (Γ, Γ′) sera le maximum de AA′ quand c varie de $-\infty$ à $+\infty$. Ce maximum est sûrement atteint, à cause de la continuité des courbes Γ et Γ′.

---

dire que (L, L′) tend vers zéro est plus délicate. Prenons par exemple pour X une variable aléatoire pouvant prendre les valeurs 2, 4, ..., $2n$, toutes ces valeurs étant également probables, et pour Y une variable admettant de même n valeurs possibles et également probables 1, 3, ..., $2n-1$; quelle que soit la corrélation entre X et Y, on a sûrement $|X - Y| \geqq 1$, et (L, L′) ne tend pas vers zéro pour n infini. En prenant au contraire les définitions de la distance que nous indiquons plus loin, on trouverait que la distance des deux mêmes lois tend vers zéro.

Pour justifier cette définition, il suffit d'observer que l'inégalité triangulaire (2) est une conséquence de l'inégalité

$$(3) \qquad AA' \leqq AA'' + A'A'',$$

écrite pour la valeur de c qui rend AA' maximum, donc égal à (L, L'). La droite $x + y = c$ coupant en A'' la courbe Γ'' qui correspond à L'', on a bien

$$AA'' \leqq (L, L''), \qquad A'A'' \leqq (L', L''),$$

et (2) en résulte

2° Voici maintenant une deuxième définition possible, qui comprend en fait une infinité de définitions, car elle utilise la notion de distance de deux points A et A' du plan, et il n'est pas nécessaire que cette distance soit la distance euclidienne; nous supposerons seulement qu'elle tende vers zéro en même temps que la distance euclidienne. De toute façon, la distance (A, Γ') de A à Γ' sera le minimum de AA quand A' décrit Γ', et la distance (L, L') = (Γ, Γ') sera le plus grand des deux nombres suivants : maximum de (A, Γ') quand A décrit Γ, et maximum de (A', Γ) quand A' décrit Γ'

Montrons que l'inégalité (2) est bien vérifiée. Prenons pour A un point quelconque de Γ, pour A'' le point de Γ'' le plus voisin de A (ou un de ces points, s'il y en a plusieurs), et pour A' le point de Γ' le plus voisin de A''. On a

$$(A, \Gamma') \leqq AA' \leqq AA'' + A'A'' \leqq (L, L'') + (L', L'')$$

La même borne supérieure s'appliquant à (A', Γ), quel que soit A' sur Γ', l'inégalité (2) en résulte.

Cette définition de (Γ, Γ') peut s'appliquer à des courbes quelconques, mais ne serait pas sans inconvénient. Elle conduirait, en effet, si, par exemple, une ellipse a son petit axe tres petit, à considérer l'ellipse complète, et la demi-ellipse située d'un côté déterminé du grand axe, comme deux courbes très peu différentes. M. Fréchet a donné une définition de la distance de deux courbes qui s'applique aux courbes de Jordan les plus générales et évite l'inconvénient que nous signalons. Nous n'en avons pas besoin ici, la définition qui précède suffisant pour les courbes très particulières que nous avons à considérer.

Nouveau rapprochement entre la distance de deux lois et celle de deux variables aléatoires. — Nous ne considérerons ici que les définitions de la distance (X, Y) qui ne dépendent que de la nature de la variable aléatoire $Z = |X - Y|$; alors

$$(4) \qquad (X, Y) = (o, Z).$$

Or, nous avons remarqué que, la notion de corrélation disparaissant lorsque l'une des variations cesse d'être aléatoire, il n'y a plus lieu de distinguer dans ce cas le voisinage des deux variables et celui des deux lois correspondantes. Peut-être même peut-on, L_0 et L_1 désignant respectivement les lois dont dépendent zéro et la variable non négative Z, réaliser l'égalité des deux distances dont il s'agit

$$(5) \qquad (o, Z) = (L_0, L_1)$$

Cette remarque conduit à poser les deux problèmes suivants.

Premier problème. — *Étant donnée une définition de* (L, L′) *acceptable* (c'est-à-dire vérifiant l'inégalité triangulaire), *la définition de* (X, Y) *que l'on en déduit par les formules* (4) *et* (5) *est-elle acceptable ?*

Deuxième problème. — *Étant donnée une définition acceptable de* (X, Y), *peut-on donner de* (L, L′) *une définition acceptable et se réduisant dans le cas de la distance* (L_0, L_1) *à celle qui résulte des formules* (4) *et* (5)?

Nous nous contentons ici de poser ces questions générales et d'indiquer une manière particulière de réaliser les conditions (4) et (5). Il n'y a qu'à prendre pour (A, A′), non la distance euclidienne de ces deux points, mais la somme

$$|x - x'| + |y - y'|,$$

x, y étant les coordonnées de A, et x' et y' celles de B, puis à définir (L, L′) en partant de (A, A′) comme nous l'avons fait au 2° du précédent paragraphe. On vérifie sans peine que la distance (L_0, L_1) est alors la distance du point $x = o$, $y = 1$ à la courbe Γ_1 correspondant à L_1, c'est-à-dire le minimum, ε variant de o à $+\infty$, de la somme

$$\varepsilon + \text{Pr.}[|X - Y| > \varepsilon],$$

ce qui revient au même que de dire la borne inférieure de

$$\varepsilon + \Pr[|X - Y| \geqq \varepsilon].$$

C'est précisément la première des définitions de (X, Y) indiquée dans cet ouvrage (p. 194).

L'espace des variables aléatoires. — Nous avons réservé pour la fin une remarque par laquelle il eût été plus logique de commencer; mais nous n'avons pas voulu risquer dès le début de décourager le lecteur par une notion assez abstraite et difficile à bien comprendre : si la notion de la distance de deux variables aléatoires est bien claire, celle d'espace des variables aléatoires l'est beaucoup moins. Il existe *des espaces de variables aléatoires*, mais parler de *l'espace des variables aléatoires*, sous-entendant par là qu'il s'agit d'un espace qui contienne toutes les variables aléatoires concevables, est aussi illusoire que de parler de l'ensemble de tous les ensembles concevables.

Je vais m'expliquer mieux. Lorsque nous parlons, par exemple, de l'espace des fonctions continues, nous pouvons le considérer comme préexistant, chaque point correspondant à une fonction bien déterminée, et nous ne risquons pas d'avoir jamais à considérer une fonction continue qui ne soit pas représentée par un point de cet espace. Au contraire, quelles que soient les variables aléatoires que nous avons déjà considérées, quelque grande que soit la puissance de l'ensemble qu'elles forment, rien ne nous empêche de considérer une nouvelle variable aléatoire indépendante des précédentes. *L'espace des variables aléatoires* est une construction qui n'est jamais terminée.

Observons, d'autre part, qu'on peut trouver, en aussi grand nombre que l'on veut, des variables aléatoires $X_1, X_2, \ldots$ dépendant d'une même loi de probabilité, mais indépendantes les unes des autres; il leur correspondra des points $A_1, A_2, \ldots$. Or, rien ne distingue ces variables les unes des autres; l'ordre seul dans lequel nous les avons introduites fait que la variable X_1, plutôt qu'une autre, correspond au point A_1. Il n'y a pas une correspondance préétablie.

La définition donnée par M. Fréchet (p. 192) est, d'ailleurs, parfaitement correcte. En parlant d'*une certaine catégorie d'épreuves*, il suppose arrêtée à un instant déterminé la construction de l'espace des variables aléatoires. Il m'a semblé qu'il n'était pas inutile d'indiquer plus explicitement la difficulté qu'il a ainsi écartée.

Je termine par une remarque relative au cas, seul intéressant en pratique, où l'on ne considère que des variables aléatoires qui soient fonctions d'une infinité dénombrable de variables indépendantes. On sait (*Voir* par exemple P. Lévy, *Bull. Sc. Math.*, 1931, p. 84 et suiv.; p. 87) que tous ces choix se ramènent au choix d'une seule variable t choisie entre 0 et 1 (avec une répartition uniforme de la probabilité dans cet intervalle), et toutes les variables aléatoires considérées sont des fonctions mesurables de t. On peut alors considérer l'espace de variables aléatoires que l'on veut étudier comme appliqué sur l'espace des fonctions mesurables (et prendre dans ces deux espaces des définitions de la distance qui se correspondent). Cette application, possible d'une infinité de manières, ne supprime pas la difficulté signalée, car il suffit d'adjoindre aux précédentes une nouvelle variable aléatoire indépendante des autres pour que tout soit à recommencer; le mode d'application considéré devra être remplacé par un autre.

NOTE C

ADDITIONS DIVERSES

Nous insérerons ici quelques additions inspirées par la correction des épreuves en indiquant les pages auxquelles elles se réfèrent.

ADDITION α.

Page 206. — *Propriétés de la convergence en moyenne quadratique dans l'espace* A_2.

Ne considérons dans ce qui suit que des variables aléatoires de l'espace A_2 défini page 208.

On a toujours

$$|\mathfrak{M}X| \leqq \mathfrak{M}|X| \leqq \sqrt{\mathfrak{M}X^2},$$

par suite, dans l'espace A_2, chaque variable aléatoire a une valeur moyenne finie et déterminée.

On a de même

$$(\bar{X}_n - \bar{X})^2 = [\mathfrak{M}(X_n - X)]^2 \leqq \mathfrak{M}(X_n - X)^2.$$

Donc *si* X_n *converge en moyenne quadratique vers* X *de* A_2, *la valeur moyenne de* X_n *converge toujours vers celle de* X (c'est ce qui n'a pas toujours lieu pour la convergence en probabilité : *voir* p. 176).

Soient $Y_n = X_n - \bar{X}_n$ et $Y = X - \bar{X}$, on a toujours

$$\mathfrak{M}(X_n - X)^2 = \mathfrak{M}(Y_n - Y)^2 + (\bar{X}_n - \bar{X})^2.$$

Des lors, pour que X_n converge vers X en moyenne quadratique, il faut et il suffit : 1° que la moyenne de X_n converge vers celle de X, 2° que la déviation $Y_n = X_n - \overline{X}_n$ de X_n converge en moyenne quadratique vers la déviation $Y = X - \overline{X}$ de X.

M. Cantelli a bien voulu me communiquer une condition nécessaire et suffisante pour cette seconde condition

Pour qu'une variable aléatoire Y_n dont la valeur moyenne est nulle converge en moyenne quadratique vers une variable aléatoire Y (dont la moyenne arithmétique sera, alors, nécessairement nulle), il faut et il suffit :

a, que l'écart quadratique moyen, μ_n, de Y_n converge vers celui μ de Y,

b, que le coefficient de corrélation r_n (défini p. 73) de Y_n et de Y tende vers l'unité.

(Ceci suppose toutefois $\mu \neq 0$.)

En effet, on a en vertu de la condition (IV) de la page 206

$$|(Y_n, 0) - (Y, 0)| \leqq (Y_n, Y)$$

ou

$$|\mu_n - \mu| \leqq (Y_n, Y) \tag{1}$$

et l'on a, d'autre part,

$$2\,\mathfrak{M}(Y_n Y) = \mu_n^2 + \mu^2 - \mathfrak{M}(Y_n - Y)^2 \tag{2}$$

Donc si Y_n converge en moyenne quadratique vers Y, μ_n tend vers μ d'après (1) et $\mathfrak{M}\, Y_n Y$ tend vers μ^2 d'après (2). Mais on a

$$r_n = \frac{\mathfrak{M}\, Y_n Y}{\mu \mu_n};$$

donc, si $\mu \neq 0$, r_n tend alors vers l'unité

Inversement, comme on a

$$\mathfrak{M}(Y - Y_n)^2 = \mu^2 + \mu_n^2 - 2 r_n \mu \mu_n,$$

si a et b sont vérifiés, le second membre tend vers zéro : Y_n converge vers Y en moyenne quadratique.

Addition β

Page 222. Une autre manière d'établir les points II et III de la page 212 consisterait à s'appuyer sur un lemme de M. Cramér (2, p. 7) concernant le cas plus général des sommes de variables aléatoires et d'où il suit, en particulier, que pour toute suite de nombres positifs λ_n, les deux séries

$$\sum_n \Pr\{|s - np| > \lambda_n\sqrt{2npq}\} \quad \text{et} \quad \sum_n \frac{e^{-\lambda_n^2}}{\lambda_n}$$

convergent ou divergent en même temps.

Addition γ

Page 100. Cette proposition peut être précisée de façon à borner l'ordre de grandeur de l'erreur. Une conséquence particulière d'un résultat de M. Cramér (1, p. 19) fournit en effet l'inégalité remarquable

$$|P_n(\lambda, \lambda') - I(\lambda, \lambda')| < \frac{6 \log n}{\sqrt{npq}}$$

pour tout entier $n > 1$, toute valeur de $p > 0$ et < 1 et pour tout couple de valeurs $\lambda, \lambda' (\lambda \leq \lambda')$.

LISTE BIBLIOGRAPHIQUE

DES MÉMOIRES CITÉS DANS CE PREMIER LIVRE

C. A. DELL'AGNOLA. — 1. *Intorno alle successioni di variabile casuali discontinue tendenti ad una variabile casuale limite* (*Atti R. Istit. Veneto Sc. Let.*, t. LXXXVIII, 1928, p. 40-62).

S. BANACH. — 1. *Sur les problèmes de la mesure* (*Fund. Math.*, t. IV, 1923, p. 7-23).

2. En collaboration avec KURATOWSKI. *Sur une généralisation du problème de la mesure* (*Fund. Math.*, t. XIV, 1929, p. 127-131).

S. BERNSTEIN. — 1. *Sur une modification de l'inégalité de Tchebichef* (en russe, résumé français) (*Ann. Sc. Instit. Sav. Ukraine*, Sect. *Math.* I, 1924).

2. *Théorie des probabilités* (en russe), 2e éd., Moscou, 1934.

J. BERTRAND. — 1. *Calcul des Probabilités* (*Gauthier-Villars*, Paris, 1889).

BIENAYMÉ. — 1. *Considérations à l'appui de la découverte de Laplace sur la loi des probabilités dans la méthode des moindres carrés* (*C. R. Acad. Sc.*, t. 37, 1853, p. 159-184).

E. BOREL. — 1. *Les probabilités dénombrables et leurs applications arithmétiques* (*Rend. Circ. Mat. Palermo*, t. 27, 1909, p. 247-271).

CAMP. — 1. *A new generalization of Tchebycheff's statistical inequality* (*Bull. Am. Math. Soc.*, vol. XXVIII, p. 427-432).

F. P. CANTELLI. — 1. *Intorno ad un teorema fondamentale della teoria del rischio* (*Boll. Ass. Att. It.*, 1910, p. 1-23).

2. *Intorno ad un teorema di Calcolo delle Probabilita* (*Giorn. Mat. Battag.*, vol. XLIX, 1911).

3. *La tendenza ad un limite nel senso del Calcolo delle probabilità*, (*R. C. Circ. Mat. Palermo*, t. XVI, 1916, p. 191-201).

4. *Sulla legge dei grandi numeri* (*R. A. Lincei, Mem.*, *Cl. Sc. Fis.*, vol. XI, 1916, p. 330-349).

5. *Sulla probabilità come limite della frequenza*, (*R. C. Lincei*, vol. XXVI, 1917, p. 39-45).

6. *Su due applicazioni di un teorema di G. Boole* (*R. C. Lincei*, vol. XXVI, 1917, p. 295-302).

7. *Sui confini della probabilità* (*Att. Cong. Intern. Mat.*, vol. VI, Bologna 1928, p. 47-60).

8. *Un teorema, sulle variabili casuali dipendenti che assorbe il teorema di Hattendorff nella teoria del rischio* (*Riv. Ital. d. Statist.*, Ann. I, 1929, p. 9-14).

9. *Considérations sur la convergence dans le Calcul des Probabilités* (*Ann. Sc. Inst. H. Poincaré*, vol V, 1935, p. 1-50).

10. *Sulla determinazione empirica di una legge di probabilità* (*Giorn Ist. Ital. Attuari*, Ann. IV, 1933, p. 421-424)

11. *Sulla estensione del principio delle probabilità totali ad una successione illimitata di eventi incompatibili* (*Giorn. Ist. Ital Attuari*, Anno VI 1935, p. 415-427).

G. Castelnuovo. — 1 *Calcolo delle Probabilità* (2e édition, 1928, chez Zanichelli Bologne).

H Copeland. — 1. *The theory of probability from the point of view of admissible numbers* (*Ann Math. Stat.*, vol. 3, 1932, p. 143-156).

A. Cournot. — 1. *Exposition de la théorie des chances et des probabilités* (Paris 1843, VIII-448 p.). *Voir* à ce sujet dans le numéro consacré à Cournot de la *Revue de Métaphysique et de Morale*, t XIII 1905, p. 293-543, *Les racines historiques du probabilisme* par F Mentré.

H. Cramér. — 1. *On the composition of elementary errors I* (*Skandin. Aktuarietidskrift*, vol. XI, 1928, p. 13-74)

2. *Su un teorema relativo alla legge uniforme dei grandi numeri* (*Giorn dell'Ist. Ital d Attuari*, t. V, 1934, p 1-13).

3. *Ueber eine Eigenschaft der normalen Verteilungsfunktion* (*Math. Zeitsch.*, 1936).

R. Dovaz. — 1. *Voir* Mirimanoff

G Faber. — 1 (*Sitz. Ber. Bayer. Ak.*, 1922).

P Flamant. — 1. *Sur deux fonctions attachées à une fonction sommable et leur application à la limite des intégrales de Lebesgue* (*C R*, *Acad Sci* t. 201, 1935, p. 930-932).

M. Fréchet. — 47. *Sur les fonctionnelles linéaires et sur l'intégrale de Stieltjes* (*C. R. Congrès Soc Sav*, 1913, Grenoble, p 46-54)

55. *Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait* (*Bull Soc. Math. Fr*, t. XLII, 1915, p. 248-265).

114. *Sur l'hypothèse de l'additivité des erreurs partielles* (*Bull. Sciences Math*, t. 63, 1928, p 203-212)

115. *Sur la loi de probabilité de l'écart maximum* (*Soc. Pol. Math Ann* t. 6, 1927, p. 93-116).

135. *Sur la convergence en probabilité* (*Metron*, vol. VIII, 1936, p 1-50).

136 *Le Calcul des Probabilités, deux conférences à la station radiotel Tour Eiffel* (*Bull A. F. A. S.*, 1930, p. 1-19).

137-139. *Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements* (deux notes, *R. C.*, *R. Istit. Lombardo Sc e Let.*, vol. LXIII, 1930).

138. *Nouvelles expressions de la distance de deux variables aléatoires et de la distance de deux fonctions mesurables* (*Ann. Soc. Pol. Math.*, t IX 1930, p. 45-49).

140. *Le generalizzazione delle ineguaglianze di Bienaymé* (*Giorn. dell Ist Ital d. Attuari.* 1931, Anno II, p. 22-36).

141. *A proof of the generalized second limit-theorem in the Theory of Probability, by M. Fréchet and J. Shohat* (*Trans. Am. Math. Soc.*, 1931, vol. 33. p. 533-543).

148 *Osservazioni ad una nota di R. Cultrera sul concetto di convergenza de una successione di variabili casuali* (*Giorn Ist Ital. Attuari*, Anno III 1932, p 273-276).

159. *Sur le coefficient dit de corrélation et sur la corrélation en général* (*Revue Inst Int Stat*, t. I, 1933, p. 1-8).

169 *Sur les précisions comparées de la moyenne et de la médiane* (*Journal Tch.-slov. sc. actuarielles*, Prague, vol 5, 1935, p 29-34).

176 *Sull espressione esatta di un scarto medio* (*Giorn. Ist Ital Attuari* Anno VI, 1936, p. 154-169)

177 *Généralisations du théoreme des probabilités totales* (*Fund Math* t 25, 1935, p 379-397)

183. *Sur quelques définitions possibles de l'intégrale de Stieltjes* (*The Duke Math. Journ*, vol. 2, 1936, p 383-395).

R. Frisch — 1. *Impossibilité de resserrer l'inégalité de Markoff dans le cas général* (*C R. Congr Scand* 1925, Copenhague, p 203-206), *Werke* Bd IV, p 11)

C F Gauss. — 1 *Theoria combinationis observationum*, Art 10-11

V Glivenko. — 1. *Sur les formes générales de la loi des grands nombres dans les espaces fonctionnels* (*Rendic. Acc. Lincei*, vol. VIII, 1928, p. 673-676 vol IX, 1929, p. 466-469; vol. IX, 1929, p 830-833)

3. *Sulla determinazione empirica delle leggi di probabilita* (*Giorn Ist Ital Attuari*, Anno IV, 1933, p 92-99)

4. *L'intégrale de Stieltjes* (volume publié en russe 1936)

K. G. Hagstroem — 1. *Bemerkungen zur Theorie der statistischen Funktion* (*Skandin, Aktuarietidskrift*, Bd II, 1919, p 204-213)

F Hausdorff — 1 *Grundzuge der Mengenlehre* (Veit, 1914) *

B. Jessen — 1 *The theory of integration in a space of an infinite number of dimensions* (*Acta Mat*, vol 63, 1934, p 249-323).

Ch Jordan. — 1. *Le théorème de probabilité de Poincaré généralisé au cas de plusieurs variables indépendantes* [*Acta Scientiarum Mathematicarum, Szeged*, t. VII, 1934, p 108, formule (4)]

Kamke. — 1 *Einfuhrung in die Wahrscheinlichkeitsrechnung* (*Leipzig*, 1932)

2 (*Jahresber d deutsch Math Ver.*, Bd 42, 1933).

Khintchine. — 1 En collaboration avec Kolmogoroff *Ueber Konvergenz von Reihen deren Glieder durch den Zufall bestimmt werden* (*Recueil Math. Moscou*, t 32, 1925, p. 668-677)

2. *Sur la loi forte des grands nombres* (*C. R. Acad Sc*, t 186, 1928, p 285)

3 *Sur la loi des grands nombres* (*C. R Acad Sc*, t 188, 1929, p 477-480)

4 (En russe dans le Progres des Sciences russes, 1930).

A Kolmogoroff. — 1. *Voir* Khintchine [1].

2. *Ueber die Summen durch den Zufall bestimmter unabhangiger Grossen* (*Mat. Ann.*, Bd. 99, 1928, p 309-318).

3. *Sur la loi forte des grands nombres* (*C R. Acad. Sc.*, t. 191, 1930, p. 910-911).

4. *Grundbegriffe des Wahrscheinlichkeitsrechnung* (Springer, 1933).

Kramp. — 1. *Analyse des réfractions astronomiques et terrestres* (Strasbourg et Leipzig, 1799).

C Kuratowski. — 1. *Voir* Banach (2).

S. LAPLACE. — 1. *Mémoire sur la probabilité des causes par les événements* (*Mémoires de l'Acad. R. Sc. Paris*, Savants étrangers, VI, 1774).
2. *Mémoire sur les probabilités*, 1780 (*Œuvres*, 1893, t. IX, p. 383).
3. *Mémoire sur les approximations des formules*, (*Œuvres*, t. XII, p. 301).
4. *Mémoire sur les intégrales indéfinies* (*Œuvres*, p. 357).
5. *Théorie analytique des probabilités* (*Œuvres*, t. XII, p. 309).

H. LEBESGUE. — 1. *Leçons sur l'intégration et la recherche des fonctions primitives* (Gauthier-Villars, 2e édition, 1928).

P. LÉVY. — 1. *Calcul des probabilités* (Gauthier-Villars, 1925).
2. *Sur les conditions d'application et sur la régularité des procédés de sommation des séries divergentes* (*Bull. Soc. Math.*, t. 49, 1925, § 7, 8).
3. *Sur les séries dont les termes sont des variables éventuelles indépendantes* (*Studia Mathematica*, vol. III, 1931, p. 119-155).
4. *Propriétés asymptotiques des sommes de variables aléatoires enchaînées* (*Bull. Sc. Math.*, t. 59, 1935, p. 84-96 et 109-128).
5. *Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées* (*Journ. de Math.*, t. XIV, 1935, p. 347-402).
6. *Sur la sommabilité des séries aléatoires divergentes* (*Bull. Soc. Math. France*, t. LXVIII, 1935, p. 1-35).
7. *Sur la notion de probabilité conditionnelle* (*Bull. Sc. Math.*, t. LX, 1936, p. 66-71).

P. MEDOLAGHI. — 1. *C. R. 7e Congr. Intern. Actuaires*, Vienne, 1909, vol. 1.

B. MEIDELL. — 1. *Sur un problème de calcul des probabilités* (*C. R.*, vol. 175, 1922, p. 806).
2. *Sur la probabilité des erreurs* (*C. R. Acad. Sc.*, t. 176, 1923, p. 280).

MIRIMANOFF. — 1. En collaboration avec R. DOVAZ. *Les épreuves répétées et la formule de Laplace* (*C. R. Acad. Sc.*, t. 185, 1927, p. 827).

M. VON MISES. — 1. *Fundamentalsatze der Wahrscheinlichkeitsrechnung* (*Math. Zeitschrift*, Bd. 4, 1919, p. 1-57).
2. *Ueber einige Abschatzungen von Erwartungswerten* (*Journ. f. reine u. ang. Math.*, Bd 165, 1931, p. 184-193).
3. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik* (*Wien, Deuticke*, 1931).
4. *Théorie des probabilités. Fondements et applications* (*Ann. Inst. H. Poincaré*, vol. III, 1932, p. 137-190).
5. *Fragen der Wahrscheinlichkeitsrechnung* (*Verhandl. d. Internat. Math. Kongr. Zurich*, 1932, Bd. II, p. 221-228).

DE MOIVRE. — 1. *Miscellannea Analytica, Supplementum*, 1733.
2. *Doctrine of Chance*, 1756.

SETEMASSU NARUMI. — 1. *On further inequalities with possible applications to problems in the theory of probability* (*Biometrika*, vol. XV, 1923, p. 245-253).

KARL PEARSON. — 1. *Historical note on the origin of the normal curve of errors* (*Biometrika*, 1924, vol. XVI, p. 402-404).

J. RADON. — 1. *Uber die absoluten additiven Mengenfunktionen* (*Ber. Ak. Wiss. Wien*, 1613).

J. SHOHAT. — 1. *Voir* M. FRÉCHET (141).

E. SLUTSKY. — 1. *Ueber stochastische Asymptoten und Grenzwerte* (*Metron*, vol. V, 1925, p. 1-90).

2. *Sur un critérium de la convergence stochastique des ensembles de variables éventuelles* (*C. R. Acad. Sc.*, t. 187, 1928, p. 370-373).

H. STEINHAUS. — 1. *Sur la probabilité de la convergence de séries* (*Studia Mat.* vol. 2, 1930, p. 21-39).

P. TCHEBICHEFF. — 1. *Des valeurs moyennes* (*Journal de Math.*, série 2, t. XII, 1867, p. 177-184).

2. *Sur les valeurs limites des intégrales* (*Œuvres*).

3. *Sur les valeurs limites des sommes* (*Œuvres*).

A. VILLE. — 1. *Sur les suites indifférentes* (*C. R. Acad. Sc.*, t. 202, 1936, p. 1393-1394).

2. *Sur la notion de collectif*.

A. WALD. — 1. *Sur la notion de collectif dans le Calcul des probabilités* (*C. R. Acad. Sc.*, t. 202, 1936, p. 180-182).

E. B. WILSON. — 1. *First and Second Laws of error* (*Quaterly publication of the Amer. Statist. Assoc.* 1923, p. 841-851).

WINKLER. — 1. *Sitzungsber. Wiener Akad.* Bd 53, 1866.

TABLE DES MATIÈRES

DU

PREMIER LIVRE.

Pages

Sommaire XIII

Avertissement . . . . . [illegible]

PREMIÈRE PARTIE

Généralités sur les probabilités

CHAPITRE I.

LA NOTION DE PROBABILITÉ.

Les diverses définitions de la probabilité 1
La définition classique 1
Les définitions basées sur la fréquence . . . . . [illegible]
Les collectifs . . . . . [illegible]
La définition statistique 4
Le hasard 4
Loi expérimentale du hasard 5
Définition empirique de la probabilité 5
Remarque 5
Axiomatisation 6
Importance de la catégorie d'épreuves 6
Catégories complexes 9
Répétition, fréquence 9

CHAPITRE II.

DIVERSES EXTENSIONS DU PRINCIPE DES PROBABILITÉS TOTALES.

Cas d'un nombre fini d'événements compatibles 11
Cas d'une suite infinie d'événements 16
Conclusion 21
Probabilités limites 22

Pages

Quelques inégalités 24
Théorèmes de Borel et de Cantelli 26
Événements négligeables 27

SECONDE PARTIE

Les variables aléatoires

CHAPITRE III.

VALEURS MOYENNES DES VARIABLES ALÉATOIRES

SECTION I. — *Introduction*

Définition des variables aléatoires . 29
Fonction des probabilités totales. Premières propriétés 30
Probabilité élémentaire et densité de probabilité 33
Autres propriétés des fonctions des probabilités totales 33

SECTION II. — *Valeurs moyennes*

Définitions 35
Second mode de calcul.. 38
Valeur médiane 39
Quartile 41
Valeur moyenne d'une somme 41
Valeur moyenne d'une différence. 46
Valeur moyenne d'une fonction donnée d'une variable aléatoire 47
Remarque 56
Moments 58
Écarts moyens 58
Inégalité de Gauss 62
Moments algébriques 65
Écart médian 66
Valeur moyenne du produit de deux variables aléatoires indépendantes. 66
Généralisation.. 69
Valeur moyenne du carré d'une somme 70
Généralisation de l'égalité de Bienaymé. 73
Fonction des probabilités totales de la somme et du produit de deux variables aléatoires indépendantes.. 74
Cas des variables dépendantes... 79
Points aléatoires, points moyens 82

SECTION III. — *Épreuves répétées.*

Valeurs moyennes des fréquences 83
Nouveau calcul de l'écart moyen.. 85
Cas de Poisson....... 88
Convergence uniforme de la loi binomiale... 89
Probabilité maximum. 92

Pages

Cas où l'écart réduit n'est pas borné . 93
Réciproque. 95
Probabilité que l'écart réduit soit compris entre deux limites données. 96
Uniformité de la convergence. 98
Équivalence asymptotique. 100
Cas où $\int_{\lambda}^{\lambda'} e^{-x^2}\,dx$ tend vers zéro. 101

SECTION IV — *Historique*

Fonctions génératrices, fonctions caractéristiques. 105
Digression sur la loi des erreurs d'observation.. 108
Encore un peu d'histoire 108

CHAPITRE IV.

L'INÉGALITÉ DE BIENAYMÉ ET SES GÉNÉRALISATIONS

L'inégalité de Bienaymé. 110
Démonstration du théorème de Bernoulli . 112
Démonstration du théoreme de Poisson. 113
Extensions de ces théorèmes 113
Premiere généralisation de l'inégalité de Bienaymé par les écarts moyens d'ordres quelconques 114
Nouvelle généralisation par les A-moments 118
Écarts moyens relativement à une fonction donnée. 120
Cas où l'on connaît deux écarts moyens. 122
Inégalités de Cantelli 123
Une méthode générale. 126
Moyennes conditionnées. . 128
Application à l'inégalité de M. Kolmogoroff .. 128
Une inégalité de Serge Bernstein . 131
Formules de Gauss-Winkler et de Camp.. 136
Cas de Gauss et cas de Camp, leurs généralisations 137
Formules préparatoires. 139
Calcul de M. Camp. 143
Cas de Gauss généralisé 144
Cas de Camp généralisé. 148
Les nouvelles formules. 153
Cas particuliers.. 154
Calcul numérique.. 155

CHAPITRE V.

LES DIVERS MODES DE CONVERGENCE D'UNE SUITE DE VARIABLES ALÉATOIRES

SECTION I — *Introduction*

Premier point de départ..... 158
Second point de départ.. 160
Quelques inégalités utiles... 162

SECTION II. — *Convergence « en probabilité »*

	Pages
Définition et propriétés	164
Premier critère de convergence en probabilité	167
Limite de la suite des fonctions de probabilités totales....	169
Signification de l'identité de deux fonctions des probabilités totales	173
Réciproque	174
Second critère de la convergence en probabilité	174
Convergence de la médiane	175
Comportements divers de la moyenne	176
Convergence en probabilité des fonctions continues	177
Cas de plusieurs variables	178
Valeur moyenne d'une fonction	180
Cas des écarts moyens	181
Application à une condition suffisante pour la convergence en probabilité	187
Autre forme de la condition	188

SECTION III. — *Premier espace de variables aléatoires*

Définition du nouvel espace	192
Distance de deux variables aléatoires	193
Troisième critère de convergence en probabilité	196
Condition pour qu'un ensemble de variables aléatoires soit compact (en probabilité)	197
Cas d'une suite qui converge en probabilité	198
Cas d'un ensemble quelconque de variables aléatoires	199
La condition est suffisante	200
La condition est nécessaire	201

SECTION IV. — *Convergences de diverses natures et espaces correspondants*

Convergence au sens ordinaire	202
Quasi-équivalence asymptotique	204
Convergence en moyenne quadratique	205
Application	207
Réciproque	208
Convergence en moyenne d'ordre r	208
Ensembles compacts dans l'espace A_r	209
La « f-convergence » en moyenne	211
Cas de réduction à la convergence en probabilité	212
Définition d'un écart de deux nombres aléatoires	212
Définition d'une nouvelle distance de deux variables aléatoires....	213
Nouvelle distance de deux fonctions mesurables	214

SECTION V. — *Convergence presque certaine.*

Définition	215
Théorème de M. Borel	215
Démonstration analytique	217
Démonstration géométrique	223
Convergence forte	228

	Pages
Convergence uniforme en probabilité	229
Suite extraite d'une suite donnée.	233
La convergence presque certaine n'est pas compatible avec une « distance »	234
Démonstration indirecte	235
Démonstration directe	236
Cas exceptionnel.	237
Conséquence.	238
Premier critère de convergence presque certaine	238
Théoreme de Cantelli	239
Forme initiale du théorème de Cantelli	240
Conditions suffisantes plus larges pour la convergence presque certaine.	241
Formule de Kolmogoroff	244
Application	246
Cas intermédiaire	247
Nature de la limite.	249
Suites de sommes	250
Suites comparables..	250
Suites équivalentes	251
Critere de Kolmogoroff.	253
Généralisation de Kolmogoroff	257
Théorème de Glivenko-Cantelli	260
Convergence d'une série de variables indépendantes	261
Application	262
Généralisation	264
Suites asymptotes Définition Énoncés	264
Probabilités égales à 0 ou 1 Théoreme de Kolmogoroff Théoreme de Lévy	266

SUPPLÉMENT MATHÉMATIQUE

RAPPEL DES PROPRIÉTÉS DES FONCTIONS MONOTONES

Définition	270
Décomposition.	270
Limite d'une suite	272
Continuité uniforme	273
Généralisation	274

NOTE A.

UNE PROPRIÉTÉ NOUVELLE DE LA SECONDE LOI DE LAPLACE.

Une hypothèse de M. Paul Lévy	277
La démonstration de M. Cramér	277
Lemme	278
Application du lemme à l'hypothèse de Lévy..	282

NOTE B, PAR M. PAUL LÉVY.

DISTANCE DE DEUX VARIABLES ALÉATOIRES ET DISTANCE DE DEUX LOIS DE PROBABILITÉ

Remarques préliminaires.	286
Définitions directes de la distance de deux lois.	288

Pages

Nouveau rapprochement entre la distance de deux lois et celle de deux variables aléatoires 290

L'espace des variables aléatoires 291

NOTE C

Additions diverses 293

LISTE BIBLIOGRAPHIQUE 297

TABLE DES MATIÈRES 303

RECHERCHES THÉORIQUES MODERNES

SUR

LE CALCUL DES PROBABILITÉS

SECOND LIVRE

Méthode des fonctions arbitraires.
Théorie des événements en chaîne
dans le cas d'un nombre fini d'états possibles

OUVRAGES DU MÊME AUTEUR

MATHÉMATIQUES APPLIQUÉES

L'équation de Fredholm et ses applications à la physique mathématique, en collaboration avec M. H. B. HEYWOOD, avec une Préface et une Note de M. J. HADAMARD. In-8, 165 pages (*épuisé*). Hermann, Paris, 1912.

Le calcul des probabilités à la portée de tous, en collaboration avec M. HALBWACHS. In-16, 300 pages, 18 exemples, 18 figures et 24 tableaux numériques. Dunod, Paris, 1924.

Nomographie (pratique et construction des abaques), en collaboration avec M. H. ROULLET (*Collection Colin*, n° 63, section mathématique). In-16, 208 pages, 79 figures. Colin, Paris, 2e édition, revue et corrigée, 1938.

Représentation des lois empiriques par des formules approchées, à l'usage des chimistes, des physiciens, des ingénieurs et des statisticiens, en collaboration avec M. R. ROMANN. In-8, VII-302 pages. Eyrolles, Paris, 1930.

MATHÉMATIQUES PURES

Développements en séries (*Recherches contemporaines sur la théorie des fonctions de variables réelles*, rédigées sous la direction de M. E. BOREL, 3e partie, 1912), *Encyclopédie des Sciences mathématiques pures et appliquées*; édition française, t. II, vol. 1, p. 210-241.

Les espaces abstraits et leur théorie considérée comme introduction à l'analyse générale (*Collection de monographies sur la théorie des fonctions*, dirigée par M. E. BOREL). In-8 de XI-296 pages. Gauthier-Villars, Paris, 1928.

L'arithmétique de l'infini, *premier des Exposés d'Analyse générale* publiés sous la direction de M. Maurice FRÉCHET. In-8, 37 pages. Hermann, Paris, 1933.

Recherches théoriques modernes sur le Calcul des probabilités (fascicule 3 du tome I du *Traité du Calcul des Probabilités*, par E. BOREL et divers auteurs). Premier Livre : *Généralités sur les Probabilités. Variables aléatoires.* In-8, XVI-308 pages, 1936. Second Livre : *Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles.* In-8, XVI-315 pages. Gauthier-Villars, Paris, 1938.

COURS DE LICENCE.

Leçons sur les séries trigonométriques. In-4° dactylographié de 62 pages (Collection : *Les Cours de la Sorbonne*). Tournier et Constant, Paris, 1935.

Théorie élémentaire des équations différentielles. In-4° dactylographié de 56 pages. Tournier et Constant, Paris, 1936.

Éléments de Calcul des Variations. In-4° dactylographié. Tournier et Constant, Paris, 1938 (*sous presse*).

Notice sur les travaux scientifiques de M. Maurice FRÉCHET. In-4° de 104 pages, Hermann, Paris, 1933.

BILITÉS

TRAITÉ du CALCUL des PROBABILITÉS et de ses APPLICATIONS

Par Émile BOREL

AVEC LA COLLABORATION DE

C.-V.-L. CHARLIER, R. DELTHEIL, P. DUBREIL, M. FRÉCHET, H. GALBRUN, J. HAAG, R. LAGRANGE, F. PERRIN, CH. RISSER, P. TRAYNARD, J. VILLE

TOME I

LES PRINCIPES DE LA THÉORIE DES PROBABILITES

FASCICULE III

RECHERCHES THÉORIQUES MODERNES SUR LE CALCUL DES PROBABILITÉS

SECOND LIVRE

Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles

Par Maurice FRÉCHET

PROFESSEUR DE CALCUL DIFFÉRENTIEL ET INTÉGRAL A LA FACULTÉ DES SCIENCES DE PARIS

PARIS

GAUTHIER-VILLARS, IMPRIMEUR-ÉDITEUR

LIBRAIRE DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE

55, Quai des Grands-Augustins, 55

1938

PRÉFACE.

Je me proposais d'indiquer dans ce Second Livre, le principe de la méthode des fonctions arbitraires et surtout de donner un exposé complet de la Théorie des probabilités en chaîne. Mais au fur et à mesure de la préparation de ce Livre, cette théorie se complétait et se développait de telle sorte que mon but ne pouvait plus être atteint qu'en gonflant excessivement ce volume, au risque d'en retarder considérablement l'impression.

D'accord avec M. Borel, je me suis donc décidé à réduire le problème résolu ici au cas le plus simple, mais dont l'étude est le plus achevée, celui des états possibles en nombre fini et des événements en chaîne simple et constante. Il est clair d'ailleurs que le cas des états possibles en nombre infini et des chaînes multiples et variables n'est qu'une généralisation du précédent; mais cette généralisation n'est pas automatique et présente même de nombreuses difficultés. Il faut souhaiter qu'elle puisse être un jour exposée en détail (par exemple dans le cadre tout indiqué de la Collection de Monographies sur le Calcul des Probabilités fondée récemment par M. Borel). Néanmoins les résultats déjà obtenus montrent qu'elle pourra s'effectuer en suivant dans ses grandes lignes l'étude du cas particulier qui fait l'objet de ce Livre. Celui-ci aura donc une portée plus grande qu'il ne paraît au premier abord. Et d'autre part

le lecteur qui s'intéresserait aux cas généraux ne sera pas laissé sans guide dans la foule des mémoires qui les traitent. Je puis annoncer en effet la publication prochaine, dans la Collection d'Exposés d'Analyse générale que j'ai l'honneur de diriger, de deux opuscules qui viendront combler cette lacune. Dans l'un, MM. Onicescu [2] (1) et Mihoc ont traité, entre autres, des chaînes multiples et des chaînes variables. Dans l'autre, M. Hostinský [19] étudiera le cas où les événements possibles forment une suite continue. Enfin, on trouvera de nouveaux résultats sur ces sujets dans les Thèses en cours d'impression de M. Dœblin [4] et de M. Fortet [1].

La matière de ce Second Livre, réduite comme il vient d'être expliqué, paraissait former une théorie complètement achevée au moment où j'ai entrepris de l'écrire. Et pourtant je n'ai pu introduire pendant sa rédaction tous les résultats récemment obtenus, comme, par exemple, l'extension due à M. Dœblin de la loi du logarithme itéré aux probabilités en chaîne. On trouvera celle-ci ainsi que d'autres résultats dans un opuscule rédigé par M. Dœblin [3] pour précéder sa Thèse et qui paraîtra prochainement.

Sauf les développements de cette sorte, énoncés en dernière heure, j'espère avoir pu cependant présenter ici un exposé assez complet des résultats obtenus dans le cadre précis auquel je me suis limité.

Je ne me suis pas contenté de rassembler et de mettre en ordre les importants résultats obtenus par divers auteurs. Attiré par cette théorie des chaînes — fondée par Markoff et Poincaré — quand elle était déjà en bonne voie de développe-

---

(1) Les chiffres entre crochets renvoient à l'index bibliographique du présent Livre, p. 301.

ment, je crois avoir réussi à la perfectionner sur nombre de points. Et j'ai pu, en particulier, résoudre complètement certains problèmes dont l'étude avait été amorcée auparavant sans avoir abouti à des résultats définitifs.

Dans la Préface du Premier Livre, j'avais fait observer que des travaux d'une grande importance qui ne s'ajustaient pas au plan assigné dès ce moment au Premier et au Second Livre, pourraient trouver leur place dans le fascicule de ce Traité prévu sous le titre « Compléments divers ». Depuis lors un grand nombre des résultats auxquels je pensais ont été exposés dans deux ouvrages importants publiés en 1937 par M. Paul Lévy [1] et par M. Harald Cramér [1]. Dans ces conditions, la publication de ce volume de « Compléments divers » paraissait moins nécessaire et sa suppression, qui a été décidée par M. Borel, permettra de hâter l'achèvement du Traité.

Dans ce Second Livre, de même que dans le Premier Livre, j'ai voulu pouvoir être lu par les nombreuses personnes qui utilisent la Théorie des probabilités sans être des mathématiciens de profession. A cet effet, je n'ai pas craint d'entrer dans les détails (sans renoncer à mettre l'essentiel en évidence) et d'autre part j'ai reporté à des notes complémentaires placées à la fin du volume, le rappel de propriétés purement mathématiques qui m'étaient nécessaires sans être classiques.

J'ai été soutenu dans la correction des épreuves par les avis scientifiques précieux de MM. Hostinský, Kolmogoroff, de Misès, Onicescu, ainsi que par les nombreuses et utiles remarques de M. Dœblin. Je leur en exprime ici tous mes remerciements.

J'ai eu encore une fois à me louer de la libéralité avec laquelle la maison Gauthier-Villars a bien voulu accueillir les nombreuses retouches qui m'ont permis de tenir compte, même au cours de la correction des épreuves (par exemple, dans

les Suppléments aux Notes A, C, D), de résultats des plus récents.

On trouvera immédiatement après cette préface un sommaire donnant une idée d'ensemble du contenu de cet Ouvrage. Celui-ci se termine par un index bibliographique, un index alphabétique, un index des notations et une table des matières détaillée.

MAURICE FRECHET

15 août 1937

SECOND LIVRE

Méthode des fonctions arbitraires.
Théorie des événements en chaîne
dans le cas d'un nombre fini d'états possibles.

SOMMAIRE DU SECOND LIVRE

Pages

PRÉFACE DU SECOND LIVRE. SOMMAIRE. LA RÉGULARISATION DES PROBABILITÉS ... VII

PREMIÈRE PARTIE

MÉTHODE DES FONCTIONS ARBITRAIRES ... 3

SECONDE PARTIE

THÉORIE DES PROBABILITÉS EN CHAINE

CHAPITRE I. — **Généralités** ... 11

CHAPITRE II. — **Nombre fini d'états possibles**

Section I. *Cas de Markoff (suite discrète d'épreuves)*

Première méthode. Étude du cas régulier par le principe de moyenne ... 22

Seconde méthode. Compléments à l'étude du cas régulier et étude du cas singulier au moyen de l'expression de $P^{(n)}_{jk}$ en fonction de n ... 102

Troisième méthode. Méthode directe ... 187

Section II. *Cas d'une suite continue d'épreuves*

Première méthode. Réduction à un système d'équations différentielles ... 204

Seconde et troisième méthodes. Solution sous forme d'une série d'intégrales multiples d'ordres croissants ... 214

Quatrième méthode. Solution en termes finis ... 219

COMPLÉMENTS DE MATHÉMATIQUES PURES

1° Quatre notes (A, B, C, D) sur les systèmes d'équations linéaires aux différences finies du premier ordre à coefficients constants ... 256

2° Notes diverses (E, F) ... 291

INDEX BIBLIOGRAPHIQUE ... 301

INDEX ALPHABÉTIQUE ... 305

INDEX DES NOTATIONS ... 307

TABLE DES MATIÈRES ... 309

RECHERCHES THÉORIQUES MODERNES

SUR LE

CALCUL DES PROBABILITÉS

SECOND LIVRE

PREMIÈRE PARTIE.

LA MÉTHODE DES FONCTIONS ARBITRAIRES.

En inventant la méthode des fonctions arbitraires, Poincaré a apporté l'une des contributions les plus importantes aux progrès modernes de la Théorie des Probabilités. Pour rendre un peu plus claire la signification que nous avons donnée plus haut de cette méthode, nous en traiterons un exemple.

CHAPITRE UNIQUE.

Problème de la roulette. — Un cercle mobile autour de son centre, la roulette, porte sur sa circonférence des arcs alternativement rouges et noirs. Les arcs rouges ont la même longueur R, les arcs noirs, la longueur N, il y en a n de chaque couleur. Une impulsion donnée au hasard fait arrêter devant un repere fixe un arc rouge ou un arc noir. Il s'agit de chercher la probabilité p_n que ce soit une division rouge, ou tout au moins d'étudier le comportement de p_n pour n tres grand.

Sans rien supposer de plus, sans rien connaître sur la façon dont les impulsions sont données, il est impossible de résoudre le probleme. Poincaré l'a posé et résolu en imposant deux conditions bien naturelles, dont la nécessité ne pourrait etre démontrée, mais dont les conséquences s'accordent avec l'expérience

Avant lui, on se serait contenté de dire : si l'impulsion est assez vive, deux arcs égaux, unicolores ou bariolés, d'un seul tenant ou non, ont la même probabilité de s'arrêter devant le repère.

Par conséquent, la probabilité p_n que ce soit un point de l'ensemble des arcs rouges et la probabilité complémentaire $1-p_n$ sont proportionnelles aux longueurs correspondantes nR, nN. De

$$\frac{p_n}{n\mathrm{R}} = \frac{1-p_n}{n\mathrm{N}},$$

on tire :

$$p_n = \frac{\mathrm{R}}{\mathrm{R}+\mathrm{N}},$$

dont la valeur est ainsi completement déterminée. Cette valeur est indépendante de n et est égale au rapport de la longueur rouge totale à la longueur de la circonférence.

L'hypothèse faite est assez exacte tant qu'on exclut les petites impulsions. Elle devient évidemment tout à fait fausse dans le cas extrême où les impulsions sont petites et où les longueurs N et R sont

assez grandes. On peut donc se demander ce qu'il en est dans les cas intermédiaires.

Pour y arriver, Poincaré a fait aussi une hypothèse, mais une hypothèse beaucoup plus générale que la précédente. Il suppose que les probabilités que deux arcs de longueurs α, β s'arrêtent devant le repère sont encore proportionnelles à α, β, mais seulement quand les deux arcs sont petits et voisins. Si on leur suppose la même origine, l'arc dont on a tourné la roulette devra être compris entre θ et $\theta+\alpha$ dans le premier cas, θ et $\theta+\beta$ dans le second. Les deux probabilités seront de la forme $k\alpha$ et $k\beta$, si α et β sont petits. Si l'on suppose en outre que le coefficient k ne dépend que de l'arc de rotation, cela revient à admettre qu'il y a une probabilité élémentaire de la forme $f(\theta)\,d\theta$ pour que l'arc de rotation soit compris entre θ et $\theta+d\theta$ (1), ou, plus précisément, on suppose que la probabilite pour que l'arc de rotation soit compris entre θ_1 et θ_2 ($\theta_1 \leqq \theta_2$) est $\int_{\theta_1}^{\theta_2} f(\theta)\,d\theta$ quels que soient θ_1 et θ_2.

Nous avons tenu à bien marquer que malgré la généralité de $f(\theta)$, il y a là une hypothèse nettement restrictive. Mais ce n'est pas tout. Poincaré ([7], p. 148) (2), suppose encore que $f(\theta)$ est une fonction dérivable M. Borel a obtenu ensuite le même résultat en supposant seulement $f(\theta)$ continue. Nous allons même nous contenter (Fréchet, [1]) d'admettre que $f(\theta)$ est intégrable. C'est la supposition la plus générale qu'on puisse faire, car cette supposition est implicitement contenue dans la définition même de $f(\theta)$. [Bien entendu, cette même définition implique aussi que $f(\theta)$ reste $\geqq 0$ et que l'on a $\int_0^{+} f(\theta)\,d\theta = 1$.]

Aucune complication de la démonstration ne résulte de cette plus grande généralité (3). Représentons par une courbe la fonction $y=f(\theta)$.

(1) Il est bien clair qu'il n'y a pas là une imprécision de langage propre au Calcul des probabilités, mais une façon de parler à qui l'on peut donner exactement et absolument la même précision que celle qu'autorisent les notions d'Analyse qu'on emploie. Cette précision varie d'une part avec les exigences de l'auteur, d'autre part avec les progrès de l'Analyse Certains raisonnements considérés comme rigoureux dans l'Analyse d'aujourd'hui ne le seront sans doute plus dans 50 ans en Calcul des Probabilités comme en Analyse.

(2) Voir p. 301, pour les références bibliographiques

(3) Toutefois nous avons supposé, pour simplifier, $f(\theta)$ bornée et intégrable au sens de Riemann

Chaque arc rouge de la roulette correspond a un segment de longueur R, chaque arc noir à un segment de longueur N. La probabilité p_n cherchée est la somme des aires des rectangles curvilignes limités par la courbe $y=f(\theta)$, par Ox et par des parallèles à Oy et ayant pour bases des segments rouges.

Soient M_k et m_k les bornes de $f(\theta)$ dans un de ces rectangles rouges, M'_k, m'_k les bornes dans un des rectangles noirs. On aura

$$\sum_k m_k R \leqq p_n \leqq \Sigma M_k R,$$

et en appelant S l'aire totale $\int_0^c f(\theta)\,d\theta = 1$,

$$\Sigma(m_k R + m'_k N) \leqq S \leqq \Sigma(M_k R + M'_k N)$$

Soit y_k l'ordonnée de $f(\theta)$ commune à un rectangle rouge et à un rectangle noir consécutifs, et

$$p'_n = \Sigma y_k R, \qquad S' = \Sigma y_k(R+N)$$

On a

$$|p_n - p'_n| \leqq \Sigma(M_k - m_k)R,$$

$$|S - S'| \leqq \Sigma(M_k - m_k)R + \Sigma(M'_k - m'_k)N,$$

$$p'_n = S'\frac{R}{R+N}$$

Supposons maintenant qu'on considère une suite de roulettes où le nombre total $2n$ des divisions noires et rouges augmente, mais où leurs grandeurs variables R, N restent dans un rapport $\frac{R}{N}$ constant.

Or, si la fonction $f(\theta)$ est intégrable, les sommes de Darboux tendent vers la même limite $S=1$, quand n tend vers l'infini. Donc

$$\Sigma(M_k - m_k)R + \Sigma(M'_k - m'_k)N$$

tend vers zéro. Il en est, par suite, de même de $\Sigma(M_k - m_k)R$. Finalement S' tend vers $S=1$, donc p'_n tend vers

$$S\frac{R}{R+N} = \frac{R}{R+N},$$

et $p_n - p'_n$ tendant vers zéro, on a

$$\lim_{n\to\infty} p_n = \frac{R}{R+N}.$$

Ce rapport qui, dans la grossière hypothèse primitivement adoptée, était une valeur de p_n indépendante de n, est devenue maintenant une limite de p_n lorsque n croît indéfiniment et une limite indépendante de la distribution initiale des probabilités définie par la fonction arbitraire $f(\theta)$. Ainsi quand n n'est pas très grand, p_n peut être notablement différent de la limite $\frac{R}{R+N}$. C'est seulement quand les divisions sont en grand nombre que le résultat obtenu en première approximation est voisin de sa valeur réelle. C'est là un premier exemple d'application de la méthode des fonctions arbitraires.

Interprétation de l'hypothèse. — L'intérêt qu'il y a à n'admettre sur $f(\theta)$ que son intégrabilité, c'est de mettre en évidence que l'hypothèse qu'on fait ne porte pas en réalité sur la fonction $f(\theta)$ mais sur la forme qu'on attribue à la répartition des probabilités qui sert de point de départ : à savoir qu'il existe une « densité » de probabilité. La fonction $f(\theta)$ est alors aussi arbitraire que le permet la forme adoptée.

Poincaré a donné dans son livre d'autres exemples. M. Hostinský ([1], [2], [3]) en a étudié aussi d'autres dans plusieurs de ses publications où il a justement insisté sur la « méthode des fonctions arbitraires ».

Méthode générale. — L'exemple de la roulette suffira toutefois pour montrer en quoi consiste la méthode. Une certaine probabilité géométrique est répartie dans un espace et dépend d'un nombre n de divisions. Pour déterminer cette répartition, au lieu d'en admettre *a priori* une uniformité qui d'ailleurs n'est généralement pas conforme aux faits, on se contente de donner un nom aux fonctions qui définissent la répartition, et l'on en cherche la limite quand n croît. On part donc de *fonctions arbitraires* et l'on arrive à une répartition déterminée. En fait, il faudra généralement assujettir ces fonctions arbitraires à des conditions de continuité, d'intégrabilité, etc., pour pouvoir établir l'existence d'une limite. Mais ce seront encore des conditions très larges.

Cette méthode venait à son heure, après une période où l'on avait admis trop facilement l'uniformité de la répartition des probabilités dans la résolution des problèmes de probabilités géométriques. Il

paraît nécessaire de réagir contre une telle tendance. Nous avons donné plusieurs exemples simples (Fréchet et Halbwachs, [76], p. 56) pour montrer que l'énoncé abstrait d'un problème géométrique ne suffit pas pour déterminer les probabilités géométriques correspondantes. Il faut pour cela *que le mode concret de réalisation des épreuves* soit aussi donné. C'est la conclusion qui ressortait déjà de l'examen du célèbre paradoxe de Bertrand. (Voir Fréchet et Halbwachs, [76], p. 65.)

Explication du hasard. — M. Hostinský [4] a précisé comment la méthode des fonctions arbitraires se rattache à la première partie de l'explication du hasard donnée par Poincaré et rappelée page xv.

M. Hostinský dit d'abord : « Supposons que dans un problème physique, nous connaissions approximativement l'état initial d'un système donné. Si nous pouvons prévoir la configuration actuelle du système avec la même approximation, nous disons que les phénomènes peuvent être prévus, qu'ils sont régis par des lois. Mais si une petite erreur sur les conditions initiales produit une erreur énorme sur la configuration finale, la prévision devient impossible et nous avons les phénomènes fortuits » M Hostinský donne alors en exemple le jet d'un dé. Les données initiales sont ici la position du corps et la distribution de ses vitesses. Il suffit de 12 paramètres pour les déterminer. Chaque état initial E correspond à un point A de l'espace à 12 dimensions. Connaissant ce point, le mouvement du dé est déterminé. Par suite, le nombre des points 1, 2, . . ., 6 qui sera finalement lu, est aussi déterminé par A. Chaque nombre x de points obtenus correspond à un certain ensemble R_x de points de l'espace E_{12} à 12 dimensions. Mais il est clair qu'une petite modification de l'état initial peut suffire pour changer d'une ou plusieurs unités ce nombre x de points. De sorte qu'au voisinage d'un point A quelconque de l'espace E_{12}, il y a des points appartenant à $R_1, R_2, \ldots, R_6$. Autrement dit, chacun des ensembles R_x est composé d'un nombre extrêmement grand de domaines partiels séparés les uns des autres. On est dans la même situation que pour l'ensemble des bandes rouges, par exemple, de la roulette On conçoit alors qu'il est peut-être possible de prouver que le nombre de ces domaines partiels est assez grand et leur répartition autour de chaque point assez uniforme pour conduire à des probabilités simples et déterminées de l'apparition de chaque nombre x

de points. Ce point de vue a été mis en œuvre dans des recherches récentes de MM. E. Hopf [1, 2], Copeland [1], de Misés [2]. C'est là un problème de mécanique qu'il serait intéressant d'étudier. M. Hostinský fait observer que, vraisemblablement, tout problème de probabilité peut être envisagé d'une façon analogue.

SECONDE PARTIE.

THÉORIE DES PROBABILITÉS EN CHAINE.

CHAPITRE I.

GÉNÉRALITÉS.

Le principe ergodique. — Nous allons examiner successivement dans ce Chapitre plusieurs problemes concrets portant sur des phénomènes de natures très différentes. Mais leurs solutions dépendent toutes du Calcul des Probabilités et, de plus, leurs traductions mathématiques sont tres analogues, sont justiciables des mêmes méthodes et peuvent être considérées comme des cas particuliers d'un même problème mathématique plus général que nous examinerons ensuite.

Dans tous ces problèmes, on se propose de vérifier un principe général, qui, énoncé d'abord par Boltzmann, sous le nom de *Principe ergodique*, sous une forme applicable seulement à la théorie cinétique des gaz, a vu récemment son sens et son champ d'application de plus en plus élargis. Dans sa forme nouvelle, à préciser dans chaque cas particulier, le « Principe ergodique » proclame que l'effet capricieux d'une opération dépendant du hasard, se trouve régularisé de plus en plus par une répétition suffisante de cette opération. Toutefois, une application brutale de ce principe — qui souffre des exceptions — conduirait à de graves erreurs. De sorte que *le problème à résoudre est plutôt de préciser* dans chaque cas particulier les circonstances — d'ailleurs toujours très générales — dans lesquelles le principe est vérifié, ou, ce qui revient au même, *le cas singulier où il se trouve en défaut.* Un autre problème, à la fois moins important et plus difficile, n'a été abordé que tout récemment, celui de préciser exactement le mode, la rapidité et même l'expression analytique de la régularisation.

Au point de vue mathématique, on peut dire que le premier problème consiste à montrer qu'une certaine probabilité dépendant de

l'état initial et de l'état final, de plusieurs variables et du nombre n des opérations, tend, lorsque n croît indéfiniment vers une probabilité limite indépendante de l'état initial. C'est en ce sens que la distribution initiale *arbitraire* des probabilités se trouve régularisée. On voit alors qu'il y a un lien étroit entre la question actuelle et la « Méthode des fonctions arbitraires » qui est une des grandes contributions de Poincaré à la Théorie des Probabilités. On ne s'étonnera donc pas que Poincaré ait été le premier à avoir aperçu la portée générale des problèmes concrets que nous allons aborder et sur lesquels il est plusieurs fois revenu

Les applications conduisent aussi à étudier le comportement quand le nombre des épreuves croît, des variables aléatoires liées aux probabilités qu'on vient de considérer.

On pourrait penser à traiter d'abord le problème le plus général et à en déduire les solutions des problèmes particuliers. Mais, outre que chacun d'eux a son intérêt propre, le problème général se présente sous une forme assez abstraite et assez complexe, de sorte que son énoncé et sa solution se saisiront plus facilement quand nous aurons d'abord examiné les cas particuliers. Et, d'ailleurs, plusieurs généralisations sont possibles.

Melange des urnes. — Le problème général que nous avons posé s'est présenté, presque dès l'origine du Calcul des Probabilités, mais sous la forme d'un problème particulier.

Bernoulli avait posé le problème suivant. On considère deux urnes U, V contenant chacune un certain nombre de boules. On tire au hasard simultanément une boule de U et une boule de V, et l'on place la premiere dans V, la seconde dans U. On répete un certain nombre n de fois cette opération. Les boules sont, les unes noires, les autres blanches, de sorte que la composition de chaque urne peut être modifiée à chaque opération. Il y a un nombre fini de compositions possibles $C_0, C_1, \ldots, C_r$ de l'ensemble des deux urnes. Et, pour une composition initiale C_h, il y a une probabilité déterminée $P_{hk}^{(n)}$ pour qu'on obtienne, après la $n^{\text{ième}}$ opération, la composition C_k. Il s'agit d'étudier ce que devient $P_{hk}^{(n)}$ lorsque n croît indéfiniment et de voir dans quelle mesure le résultat d'un très grand nombre d'opérations dépend de la composition initiale. P. S. Laplace [1] a étudié ce problème dans le cas particulier

où les deux urnes contiennent le même nombre de boules. où ce nombre est très grand et où, dans l'ensemble des deux urnes, il y a autant de boules noires et de boules blanches ([1]). Et il a démontré que la différence des valeurs moyennes du nombre de boules blanches et de boules noires dans la première urne, après la n^{ieme} opération tend vers zéro, c'est-à-dire que la composition se régularise et tend vers une distribution des boules noires et blanches qui est semblable dans les deux urnes.

La démonstration de Laplace ne concerne qu'un cas particulier et manque de rigueur. Son résultat a été établi rigoureusement beaucoup plus tard par A. A. Markoff [8, 10, 11] par une méthode toute différente, basée sur un principe de moyenne que nous aurons à employer constamment par la suite. Cependant, la méthode de Laplace est intéressante parce qu'elle le conduit à une équation aux dérivées partielles qui est d'un type retrouvé indépendamment plus tard par Smoluchowski dans le problème si différent, en apparence, de la diffusion D'autre part, la théorie de la spéculation a conduit M. Bachelier [1, 2, 3] à un autre type d'équations aux dérivées partielles. Et les équations aux dérivées partielles de Laplace, de Smoluchowski et de M. Bachelier, issues d'origines concretes aussi différentes, sont des cas particuliers d'une équation aux dérivées partielles obtenue récemment par A. Kolmogoroff [1] au cours de l'étude d'un probleme englobant ces différents problemes particuliers.

Nous reviendrons ailleurs sur ces diverses formules. Nous allons provisoirement nous contenter de ramener le problème du mélange des urnes à sa forme mathématique que nous résoudrons un peu plus loin (p. 34, 49, 69, 75 et 93) par la méthode de Markoff

Désignons par B, N les nombres de boules blanches et noires contenues dans l'ensemble des deux urnes, par u et v les nombres respectifs de boules contenues dans l'urne U et dans l'urne V On a

$$B + N = u + v.$$

Soit b, le nombre de boules blanches contenues dans U apres un nombre déterminé d'opérations. La connaissance de b détermine la composition des deux urnes : U contiendra b blanches, $u - b$ noires,

---

([1]) A vrai dire, il a aussi considéré des cas où ces hypothèses ne sont pas toutes remplies.

V contiendra $B - b$ blanches et $N - u + b = v - B + b$ noires. De sorte qu'on désignera sans ambiguïté la composition des urnes par C_b, b étant le nombre de boules blanches de U dans cette composition. D'ailleurs, si, par exemple, $N \geqq u$, c'est-à-dire $v \geqq B$, b peut prendre l'une quelconque des valeurs 0, 1, 2, ..., r, r étant le plus petit des deux nombres B et u (Si l'on avait $N < u$, c'est-à-dire $v < B$, il faudrait que l'on eût $b \geqq u - N > 0$. Alors on définirait la composition par le nombre des boules noires de U.)

Si l'on part maintenant de la composition initiale C_h, on voit que la probabilité d'obtenir la composition C_k après $m + n$ opérations sera la somme des probabilités ϖ_j d'obtenir C_k en passant à la $n^{\text{ième}}$ opération par la composition C_j.

Or, ϖ_j est, d'après le théorème des probabilités composées, le produit de la probabilité $P_{hj}^{(n)}$ de passer de C_h à C_j par n opérations et de la probabilité $P_{jk}^{(m)}$ de passer de C_j à C_k par m opérations. On a donc la formule fondamentale que nous rencontrerons à nouveau [1] :

$$(I) \qquad P_{hk}^{(m+n)} = \sum_{j=0}^{j=r} P_{hj}^{(n)} P_{jk}^{(m)}$$

En particulier, si l'on pose $p_{ij} = P_{ij}^{(1)}$, on a

$$(I_1) \qquad P_{hk}^{(n)} = \sum_{j=0}^{j=r} p_{hj} P_{jk}^{(n-1)} = \sum_j P_{hj}^{(n-1)} p_{jk}$$

De sorte que si les p_{ij} sont connus, on aura successivement les $P_{ij}^{(2)}$, $P_{ij}^{(3)}$, ... au moyen de cette relation de récurrence.

Il faut donc, pour calculer cette suite, déterminer les p_{ij}. Nous observerons qu'à chaque opération, le nombre des boules blanches de U ne peut varier de plus d'une unité. Donc

$$p_{ij} = 0 \quad \text{si} \quad |i - j| > 1$$

Observons maintenant qu'en m opérations, le nombre des boules blanches de U peut changer de m unités au plus. Donc

$$P_{ij}^{(m)} \begin{cases} = 0 & \text{si} \quad |i - j| > m, \\ \geqq 0 & \text{si} \quad |i - j| \leqq m \end{cases}$$

---

[1] A titre de moyen mnémotechnique, nous désignerons cette formule d'itération par la lettre I.

D'ailleurs, on a toujours $|i-j| \leqq r$. Il est donc déjà vraisemblable (et ce point sera élucidé plus loin, p. 33), qu'en général, pour $m \geqq r$, tous les $P_{ij}^{(m)}$ seront positifs. Il y a là une première indication d'une tendance à la régularisation.

Mais si nous voulons calculer les $P_{hk}^{(n)}$, il faut d'abord calculer effectivement ceux des p_{ij} qui ne sont pas nuls. Dans la composition C_i, la probabilité de tirer une boule blanche de U est $\frac{i}{u}$, celle de tirer une boule blanche de V est $\frac{B-i}{v}$. Pour augmenter le nombre de boules blanches de U, il faut tirer une noire de U, une blanche de V et les permuter. Donc,

$$p_{i,i+1} = \left(1-\frac{i}{u}\right)\left(\frac{B-i}{v}\right);$$

de même,

$$p_{i,i-1} = \frac{i}{u}\left(1-\frac{B-i}{v}\right),$$

et enfin,

$$p_{ii} = \frac{i}{u}\,\frac{B-i}{v} + \left(1-\frac{i}{u}\right)\left(1-\frac{B-i}{v}\right).$$

On aura évidemment,

$$1 = p_{i,i-1} + p_{ii} + p_{i,i+1},$$

ou ici

$$(\mathrm{T}_1) \qquad \sum_{j=0}^{j=r} p_{ij} = 1$$

Au moyen de ces formules, et de la formule d'itération (1), il est — au moins théoriquement — possible de calculer de proche en proche les $P_{hk}^{(n)}$.

Bien entendu, puisque les $P_{hk}^{(n)}$ sont des probabilités, on a

$$(\mathrm{P}) \qquad P_{hk}^{(n)} \geqq 0 \quad (^1)$$

Il est utile, pour la suite, de voir quels sont ceux des p_{ij} qui sont sûrement $\neq 0$. Supposant encore $v \geqq B$, l'expression de $p_{i,i+1}$ n'a de

---

[1] A titre de moyen mnémotechnique, nous désignerons cette relation (qui interviendra fréquemment par la suite) par l'initiale P du mot positif, et de même la relation plus haut par l'initiale T du dernier mot de l'expression · application du théorème des probabilités totales.

sens que si $0 \leqq i < i+1 \leqq r \leqq \left\{ \begin{matrix} B \\ u \end{matrix} \right.$. On doit donc y supposer $i<B$, $i<u$, et par suite, $p_{i,i+1} \geqq \frac{1}{uv} > 0$. De même, $p_{i,i-1} \geqq \frac{1}{uv} > 0$. Par contre, pour p_{ii}, i peut prendre toutes les valeurs de 0 à r. Il ne peut être nul que si $i(B-i)=0$ et $(u-i)[(v-B)+i]=0$. Si $i=0$, cela exige que $v-B$ supposé jusqu'ici $\geqq 0$ soit nul. Si $i=B$, cela exige que $u=B$. D'ailleurs, si $B=u$ ou v, N sera égal à v ou u.

Donc, en dehors du cas où le nombre de boules d'une même couleur est égal au nombre de boules d'une même urne, tous les p_{ij} sont $\neq 0$ pour $|i-j| \leqq 1$.

Dans les cas exceptionnels : si l'on a, par exemple, $B=u\neq v$, alors, en supposant encore $v \geqq B$, on a $B=u<v$ et $p_{rr}=0$ est la seule égalité en opposition avec la règle du cas général. Enfin, si, comme dans le cas de Laplace, $B=u=v\,(=N)$, alors $p_{00}=0$, $p_{rr}=0$ sont les deux seules égalités exceptionnelles.

Remarque. — On pourrait se demander s'il est permis de permuter les indices dans l'égalité (T_1) ci-dessus. En faisant le calcul, on trouve que

$$\sum_j p_{ji} = p_{i-1,i} + p_{ii} + p_{i+1,i} = 1 + \frac{u+v+2}{uv} > 1 \qquad \text{pour} \quad 0<i<r$$

Battage des cartes. — Ce problème a été posé et résolu par Poincaré [7, p. 301]. Quand on applique le Calcul des Probabilités aux jeux de cartes, on suppose essentiellement qu'il y a pour le joueur égale probabilité de tirer une carte ou une autre. Cela implique, bien entendu, que le joueur est dans l'ignorance complète de l'ordre des cartes dans le paquet qui lui est présenté. Autrement dit, quand il se décide à tirer du paquet une carte de rang k, il faut qu'il y ait, après battage des cartes, une égale probabilité que cette carte ait une valeur ou une autre. Si le joueur connaissait la disposition du paquet avant le battage, il faut qu'il ne puisse rien conclure du rang h de la carte avant le battage concernant son rang k après le battage. La probabilité pour que h devienne k doit donc être indépendante de k et de h.

Le battage s'effectue généralement par une succession d'opérations de même nature, par exemple par «coupage». La probabilité en question

doit donc dépendre en principe du nombre n de ces opérations, désignons-la par $P_{hk}^{(n)}$. Il est facile de voir que la condition théorique d'indépendance ci-dessus n'est réalisée ni exactement, ni même d'une manière approchée pour les petites valeurs de n, et en particulier pour $n = 1$.

Pourtant, en fait, un nombre assez faible d'opérations suffit généralement à uniformiser à peu près les chances. Il s'agit d'expliquer ce résultat et même de prévoir l'ordre de grandeur du nombre d'opérations nécessaires pour une régularisation suffisante. Il y a même lieu de rechercher si certaines opérations ne doivent pas être rejetées comme ne conduisant pas à la régularisation désirée.

Poincaré a résolu le problème d'une manière ingénieuse, mais exigeant la connaissance des nombres complexes à n parties.

Plusieurs auteurs en ont cherché une démonstration plus directe.

Il est clair qu'on a, comme dans le mélange des urnes, et par un raisonnement analogue,

$$(I) \qquad P_{ik}^{(m+n)} = \sum_{j=1}^{j=r} P_{ij}^{(m)} P_{jk}^{(n)},$$

où r est le nombre des cartes (Cette formule reste applicable, même si le passage intermédiaire d'une carte du rang i à l'un des rangs j était, en raison des procédés de battage, impossible, car alors il suffirait de prendre $P_{ij}^{(m)} = 0$) D'autre part, au bout de m battages, chaque carte en partant du rang i arrivera nécessairement à l'un des rangs j, de sorte que

$$(T) \qquad \sum_{j=1}^{j=r} P_{ij}^{(m)} = 1.$$

Enfin, on a encore la condition

$$(P) \qquad P_{ij}^{(m)} \geqq 0$$

M Urban [1, p 167] a observé qu'en vertu de l'ensemble des relations (I), (P) et (T), $P_{ik}^{(m+n)}$ est une *moyenne pondérée* des $P_{jk}^{(n)}$. Si donc $P^{(n)}$ et $p^{(n)}$ sont la plus grande et la plus petite valeur des $P_{jk}^{(n)}$ pour n donné, on aura

$$p^{(n)} \leqq p^{(m+n)} \leqq P^{(m+n)} \leqq P^{(n)}.$$

Ainsi, quand n croît, $[P^{(n)} - p^{(n)}]$ ne peut croître et, en général, les divergences des valeurs des $P_{ik}^{(n)}$ pour n donné tendent à s'effacer quand le nombre des battages s'accroît.

Cette remarque montre, en outre, que $p^{(n)}$ et $P^{(n)}$ tendent vers des limites respectives p et P. Mais elle ne nous suffit pas pour nous donner les valeurs de ces limites. Observons que si ces limites sont égales, alors $P_{ij}^{(m)}$ tend vers leur valeur commune quand m tend vers l'infini, quels que soient i et j. L'équation (T) montre qu'on aura, dans ce cas,

$$P = p = \frac{1}{r},$$

si r est le nombre des cartes. Il suffit donc de prouver que les $\left| P_{ij}^{(m)} - \frac{1}{r} \right|$ tendent vers zéro. Mais ceci n'aura certainement pas lieu dans tous les cas, comme Poincaré l'a d'ailleurs fait remarquer [7, p. 309, § 230].

M. Paul Lévy [1, p. 49] a résolu la question en affirmant que les différences $\left| P_{ij}^{(m)} - \frac{1}{r} \right|$ sont inférieures aux termes d'une progression géométrique indépendante de i et j et qui est convergente lorsque aucune des probabilités p_{ik} (qui sont les seules données et résultent du mode de battage considéré) n'est nulle (1).

M. Hadamard [1, 4], puis M. Hostinský [5, 7] ont repris la question et, tout en donnant une démonstration explicite (Voir aussi, Paul Lévy [II, p. 231]) du fait précédent, ont examiné en détail certains cas où l'un des p_{ik} est nul.

Plus tard, M. Hostinský a constaté que le problème du battage des cartes aurait pu être considéré déjà comme résolu en l'envisageant comme un cas particulier du problème général des événements en « chaînes », problème posé et résolu par Markoff en 1907 [2, 4] et que nous étudierons plus loin en détail (p. 23). La méthode de Markoff est effectivement plus simple que celle de Poincaré *pour le cas régulier* considéré par Markoff.

Par contre, il y a lieu de noter que la méthode de Poincaré, convenablement complétée, est plus puissante que celle de Markoff, qu'elle permet, par exemple, d'étudier plus profondément le cas singulier et qu'elle se prête à des extensions à d'autres théories que celle des probabilités.

Mouvements browniens discontinus. — Poincaré avait étudié le problème du battage des cartes en guise de préparation à l'étude du problème plus complexe du mélange des liquides dont il a montré l'importance et les difficultés [7, p. 320]. Bien que n'ayant aucun

(1) Nous verrons d'ailleurs plus loin (p. 32), que cette condition suffisante n'est pas nécessaire.

rapport physique, ces deux problèmes sont, au point de vue de la Théorie des Probabilités, étroitement liés. Le mélange des liquides est, au contraire, en rapport physique étroit avec le mouvement brownien étudié récemment par plusieurs auteurs, mais par des méthodes tout à fait différentes de celles esquissées par Poincaré. Dans le phénomène du mouvement brownien, on peut admettre qu'un liquide est formé de particules en mouvement rapide S'il y a deux liquides en presence, les chocs successifs des particules ont pour effet de disséminer peu à peu les molécules d'un des liquides parmi celles de l'autre. Les mouvements de ces particules sont, sans doute, déterminés par des lois physiques précises, de même que le mouvement d'une pièce de monnaie qui tombe. Mais les circonstances de toute nature qui déterminent au bout d'un temps donné les nouvelles positions des molécules sont si nombreuses — alors que de petites variations de ces circonstances peuvent grandement modifier le mouvement — qu'on peut considérer ces mouvements comme dus au hasard.

En première approximation, Lord Rayleigh [1] a étudié le mouvement d'une molécule qui se déplacerait sur une droite de telle façon qu'elle éprouve successivement des déplacements egaux a une longueur constante l, mais tantôt à droite, tantôt à gauche, les probabilités des deux directions étant égales — et par suite égales à $\frac{1}{2}$ Il a trouvé ainsi, que le déplacement résultant d'un tres grand nombre n de déplacements de même longueur l, est tel que la valeur moyenne de son carré est asymptotiquement équivalente à nl^2 en supposant illimitée la droite considérée. En réalité, elle est exactement égale à nl^2. Nous avons établi ce résultat (Fréchet, [149], p. 241) dans un cas plus étendu

Dans le cas de Lord Rayleigh, la position initiale A_0 de la molécule étant fixée, les positions possibles, séparées de la première par un nombre entier d'intervalles de longueur l forment une suite discontinue A_0, A_1, A_{-1}, A_2, A_{-2}, ... Pour aborder la seconde approximation, on peut supposer que les positions possibles d'une molécule, sans être quelconques, forment une suite discontinue de points B_1, B_2, B_3, ... donnés, mais placés en des situations quelconques, c'est-à-dire que ces points peuvent n'être ni équidistants, ni alignés.

Chaque molécule éprouve une suite de déplacements dus au hasard

qui ne peuvent la faire passer que de l'une des positions précédentes à une autre, et l'on suppose qu'il y a une probabilité déterminée p_{ik} pour passer de B_i à B_k en un seul déplacement. On appelle encore $P_{ik}^{(n)}$ la probabilité pour que le même déplacement total de B_i à B_k soit le résultat de n déplacements successifs. Il s'agit d'étudier le comportement de $P_{ik}^{(n)}$ quand n croît. Ce problème n'est qu'une des formes du problème des chaînes de Markoff dont nous allons maintenant préciser l'énoncé général.

Probleme général des chaînes — Les problèmes que nous venons de rappeler : mélange des urnes, battage de cartes, mouvements browniens, rentrent dans un type général qui a été étudié systématiquement par Markoff [1, 2, . . , 15].

Bien que Markoff et Poincaré aient été conduits aux mêmes types de problèmes, il est intéressant de noter que c'est en partant de points de vue très différents. (Nous avons d'ailleurs déjà fait observer que leurs méthodes de résolution sont également différentes.)

Le point de vue de Poincaré est très bien indiqué dans l'introduction de son livre [7]. Il ramène l'intervention du hasard à deux circonstances principales : le hasard serait dû, soit à l'action de petites causes susceptibles de produire de grands effets, soit à l'action de causes très complexes. Dans le cas du jeu de la roulette, une différence extraordinairement faible dans l'impulsion donnée, a comme effet, l'arrêt, devant la flèche, d'un arc d'une couleur différente et peut avoir comme conséquence la ruine ou la fortune du joueur. Dans le battage des cartes, au contraire, un seul coup de battage n'altère pas profondément les probabilités; un joueur qui connaissait la disposition précédente des cartes ne tirera pas entièrement au hasard une carte donnée. Mais la complexité des causes représentées ici par le grand nombre des coups de battage, mettra finalement ce même joueur dans l'impossibilité de tirer parti de ses renseignements sur la position initiale et le contraindra à s'en remettre au hasard.

L'objectif de Poincaré est alors de mettre d'abord en évidence, de chiffrer, jusqu'à un certain point, l'effet de ces deux circonstances dans les exemples simples ci-dessus. Mais il ajoute que ce genre de recherches n'a pas seulement un intérêt philosophique, que les méthodes à employer — et qu'il a indiquées — pourront être appliquées à des problèmes importants de la philosophie naturelle. On en

touchera ici quelques-uns, mais l'objet propre de ces explications étant en dehors du cadre de ce fascicule, nous renverrons sur ce point à la Note, pleine d'idées, de Poincaré sur le mélange des liquides [7, p. 320-333], ou aux fascicules du présent Traité consacrés aux *applications* du Calcul des Probabilités.

Le point de départ de Markoff est tout autre. Il s'agit avant tout, pour lui, de généraliser les propriétés classiques des événements indépendants et des variables indépendantes. Il est clair que si l'on ne sait absolument rien sur la dépendance qu'on veut introduire, une telle généralisation sera impossible, ou, tout au moins, se réduira à un très petit nombre de résultats. Il s'agissait donc d'imaginer une dépendance de nature assez lâche pour qu'on puisse lui étendre un grand nombre des résultats obtenus dans le cas de l'indépendance. Il fallait, d'autre part que cette sorte de dépendance se rencontrât effectivement dans les applications et en même temps fût assez générale pour se présenter dans des exemples de natures très diverses.

C'est ce que Markoff a su très bien réaliser au moyen de sa conception des phénomènes en chaîne. Non seulement il a su effectuer les généralisations qu'il avait en vue pour les événements « en chaîne » de l'espèce que nous allons étudier dans ce Chapitre, mais encore le cas spécial qu'il a étudié — celui où les variables et la succession des épreuves forment des suites discontinues — se prête à des extensions au cas des suites continues d'états et d'épreuves Celles-ci ont fait l'objet utile d'études postérieures.

Nous ne traiterons ici que le cas d'un nombre fini d'états possibles. L'étude du cas d'un ensemble dénombrable ou continu d'états possibles étant moins avancée sera reportée hors de ce Traité

Événements en chaînes — Si, dans le problème du mélange des urnes de la page 12, on remettait les boules dans les urnes dont elles ont été tirées, les probabilités de chaque tirage seraient independantes des résultats des épreuves antérieures. Les événements consistant dans l'obtention des divers tirages possibles seraient indépendants. En opérant, au contraire, de la façon étudiée par Laplace, ces événements sont dépendants. Mais une dépendance aléatoire peut exister sans qu'il soit besoin de faire intervenir explicitement une série d'épreuves. Par exemple, on jette un dé, et l'on obtient un nombre de points égal à n. Soient E_1 l'événement consistant en ce que $n \leqq 5$,

E_2 celui qui se présente quand *n* est pair. Aucun des deux n'est déterminé quand l'autre a lieu. Mais la probabilité de chacun d'eux diffère suivant qu'elle est relative à l'ensemble des épreuves possibles ou seulement à l'ensemble de celles des épreuves possibles où l'autre a lieu. Ce n'est pas une dépendance complète, c'est une dependance aléatoire.

La dépendance aléatoire particulière considérée par Markoff se distingue donc déjà en ce qu'*elle fait intervenir essentiellement une succession plus ou moins longue d'épreuves*.

La probabilité d'un événement lors d'une épreuve déterminée pourrait dépendre de l'ensemble des résultats des épreuves antérieures, ou, au moins, de plusieurs des épreuves antérieures, c'est-à-dire de « l'histoire » complète ou partielle, du Système considéré. On aurait affaire à une « *chaîne multiple* ». Même dans le cas d'une « chaîne simple » la probabilité de passage, en une épreuve, d'un état à un autre pourrait encore dépendre du rang de cette épreuve. Ce serait une « *chaîne variable* ». Nous laisserons aussi de côté ces généralisations, ne considérant dans ce qui suit que le cas des chaînes « *simples et constantes* ».

Au Chapitre suivant, nous étudierons dans la Section I le cas d'une « *suite discrète d'épreuves* », celui où les épreuves peuvent être numérotées, et dans la Section II le cas d'une « *suite continue d'épreuves* », les épreuves se déroulant de façon continue dans le temps.

CHAPITRE II.

NOMBRE FINI D'ÉTATS POSSIBLES.

SECTION I.

CAS DE MARKOFF SUITE DISCRÈTE D'ÉPREUVES

Hypothèse de base. — Considérons certains événements fortuits incompatibles E_1, E_2, ..., E_r, en nombre fini et dont, à chaque épreuve, l'un se produit nécessairement. Quand on sait ceux qui se sont produits aux épreuves précédant une épreuve déterminée, on ne sait pas, en général, quel est celui qui se produira, de sorte qu'il n'y a pas, en général, dépendance complète. Mais l'on suppose que la probabilité que se produise E_h dépend de ceux des événements qui se sont produits dans les épreuves précédentes. Bien que Markoff se soit aussi occupé du cas où l'on fait intervenir ainsi plusieurs des épreuves précédentes, il s'est attaché surtout au cas où la probabilité de chaque événement E_k est déterminée quand on connaît seulement celui, E_h, des événements E_1, ..., qui s'est produit à la seule épreuve précédant immédiatement l'épreuve considérée. C'est le cas des chaînes dites simples et constantes. On voit qu'on pourra représenter complètement par la notation p_{hk} la probabilité pour que l'événement E_k ait lieu quand E_h a eu lieu à l'épreuve précédente.

On peut aussi, comme le fait M. Kolmogoroff — et ce qui sera plus commode quand on arrivera aux suites continues d'épreuves — considérer $E_1, E_2, \ldots, E_r$ comme les états possibles d'un certain Système S. Alors p_{hk} sera la probabilité — bien déterminée, par hypo-

thèse, par les indices h et k — pour que le Système S passe en une seule épreuve (ou opération) de l'état E_h à l'état E_k. Dans l'exemple du mélange des urnes, l'existence de la probabilité p_{hk} a été établie par le calcul même de sa valeur. Par contre, dans le problème du battage des cartes, le fait que les « habitudes du joueur » auraient pour conséquence la détermination de p_{hk} n'est rien moins qu'évident. C'est une hypothese commode pour se rendre compte de l'effet d'une succession prolongée de coups de battage, mais il ne faudrait pas oublier qu'*il y a, dans ce problème particulier, une hypothèse assez fragile* à la base des calculs.

Par contre, si la détermination des p_{hk} est admise, la détermination de la probabilité $P_{hk}^{(n)}$ pour que le Système S passe en n opérations de l'état E_h à l'état E_k en est une conséquence certaine. Il suffit, en effet, d'appliquer les théorèmes classiques de la même façon qu'à la page 14, pour en conclure que l'on a (¹) :

$$P_{hk}^{(m+n)} = \sum_{j=1}^{j=r} P_{hj}^{(m)} P_{jk}^{(n)}, \tag{I}$$

on a encore, en particulier,

$$P_{hk}^{(n+1)} = \sum_{j=1}^{j=r} P_{hj}^{(n)} p_{jk} = \sum_{j=1}^{j=r} p_{hj} P_{jk}^{(n)} \tag{I$_1$}$$

avec

$$P_{hk}^{(1)} = p_{hk}$$

Ainsi, non seulement $P_{hk}^{(n)}$ est déterminé par la seule connaissance des indices h, k, n et, bien entendu, des valeurs des p_{hk}, mais encore on peut le calculer par récurrence au moyen de la formule (I_1).

Partant de E_h, le Système se trouvera nécessairement après n épreuves dans l'un des états E_k. Ceci montre que l'on a aussi

$$\sum_{k=1}^{k=r} P_{hk}^{(n)} = 1, \tag{T}$$

et en particulier

$$\sum_{k} p_{hk} = 1 \tag{T$_1$}$$

(¹) On observera ici une différence de notation avec le cas du mélange des urnes où il avait été utile d'admettre la valeur zéro parmi les valeurs possibles des indices. Actuellement r et non pas $r+1$ est le nombre des états possibles.

Enfin il est clair que

$$(\mathrm{P}) \qquad \mathrm{P}_{hk}^{(n)} \geqq 0,$$

et en particulier

$$(\mathrm{P}_1) \qquad p_{hk} \geqq 0.$$

Le cas considéré par Markoff, défini à la page 23, est caractérisé par les conditions (I), (T), (P).

PREMIÈRE MÉTHODE

ÉTUDE DU CAS RÉGULIER PAR LE PRINCIPE DE MOYENNE.

a — Problème du comportement de $\mathrm{P}_{hk}^{(n)}$ quand n croît

La formule (I_1) permettrait de calculer successivement les $\mathrm{P}_{hk}^{(n)}$ (Nous verrons aussi page 113 que, dans le cas général, on peut obtenir une forme explicite de $\mathrm{P}_{hk}^{(n)}$ en fonction de n. On pourra à ce moment voir plus exactement comment se comporte $\mathrm{P}_{hk}^{(n)}$ quand n croît). Mais outre que ce calcul deviendra de plus en plus difficile dans les cas de plus en plus généraux mentionnés page 22, des considérations plus simples, dues à Markoff et basées sur un « principe de moyenne », permettent de déterminer directement l'allure de ce comportement sous des hypothèses très générales.

Formulons « le principe de la moyenne » pour une valeur fixe du second indice k, relativement à l'ensemble des formules

$$\mathrm{P}_{hk}^{(n)} = \sum_j p_{hj} \mathrm{P}_{jk}^{(n-1)},$$

$$\sum_j p_{hj} = 1.$$

Il consiste à observer que, d'après ces formules jointes à la condition (P_1), $\mathrm{P}_{hk}^{(n)}$ est une moyenne pondérée des $\mathrm{P}_{jk}^{(n-1)}$ et sera par suite

compris entre leurs valeurs extrêmes. On en déduit une première conséquence.

En désignant par $P_k^{(n)}$ et $p_k^{(n)}$ le plus grand et le plus petit des nombres $P_{1k}^{(n)}, P_{2k}^{(n)}, \ldots, P_{rk}^{(n)}$, on voit immédiatement que

$$(1) \qquad p_k^{(n-1)} \leqq p_k^{(n)} \leqq P_k^{(n)} \leqq P_k^{(n-1)}$$

Les deux suites monotones et bornées

$$p_k^{(1)} \leqq p_k^{(2)} \leqq \ldots \leqq p_k^{(n)} \leqq \ldots,$$
$$P_k^{(1)} \geqq P_k^{(2)} \geqq \ldots \geqq P_k^{(n)} \geqq \ldots$$

sont donc convergentes. Elles ont donc des limites respectives p_k et P_k, et l'on a $p_k \leqq P_k$.

I

Le cas régulier. Condition nécessaire. — Il est naturel de se demander dans quel cas les limites p_k et P_k sont toutes égales deux à deux, c'est-à-dire dans quel cas chacun des $P_{hk}^{(n)}$ a une limite P_k indépendante de l'état initial et par suite indépendante du premier indice, h. C'est ce que nous appellerons, avec M. Hadamard, *le cas régulier*. Dans ce cas, chacune de ces limites serait $\geqq 0$. De plus, l'égalité $\sum_k P_{hk}^{(n)} = 1$ donnant à la limite

$$\sum_k P_k = 1,$$

l'un au moins des P_l serait positif et non nul. Et comme on aurait

$$0 < P_l = p_l = \lim_{n \succ \infty} p_l^{(n)},$$

l'un au moins des $p_l^{(n)}$, soit $p_l^{(\nu)}$, serait positif. Ainsi le cas régulier ne peut se présenter que si, parmi les tableaux $D^{(n)}$ des $P_{hk}^{(n)}$,

$$D^{(n)} \equiv \left\| \begin{matrix} P_{11}^{(n)} & \ldots & P_{r1}^{(n)} \\ \cdot & \cdot & \cdots \\ P_{1l}^{(n)} & \ldots & P_{rl}^{(n)} \\ \cdot\;\cdot & \cdots & \cdots \\ P_{1r}^{(n)} & \ldots & P_{rr}^{(n)} \end{matrix} \right\|,$$

il en existe au moins un, $D^{(\nu)}$, dont une ligne au moins a ses éléments (correspondant à un même *second* indice l, à un même état *final* E_l) différents de zéro. D'ailleurs, s'il en est ainsi, la même ligne sera positive dans tous les $D^{(n)}$ au moins à partir de $D^{(\nu)}$ (car $p_l^{(n)} \geqq p_l^{(\nu)} > 0$ pour $n \geqq \nu$).

Il y a un cas particulier du cas régulier qui est intéressant, c'est celui où, non seulement les limites P_k existent, mais où elles sont toutes positives, nous l'appellerons cas *positivement régulier* [1].

Pour qu'on soit dans ce cas, il faut évidemment, d'après ce qui précède, que l'un au moins, $D^{(n_0)}$, des tableaux $D^{(n)}$ ait tous ses éléments positifs. Et, alors, il en sera de même de tous les $D^{(n)}$, au moins à partir de $D^{(n_0)}$.

Conditions suffisantes. — Markoff a montré, au moyen d'un ingénieux raisonnement, que cette dernière condition suffit pour qu'on soit dans le cas positivement régulier.

Au moyen du même raisonnement de Markoff convenablement complété, nous allons établir l'égalité des p_k et P_k dans le cas plus général que celui de Markoff où l'*avant*-dernière condition seulement (moins stricte que la dernière et rappelée plus loin) est réalisée.

On a évidemment à majorer, quand h et h' varient, la différence

$$P_{hk}^{(n+1)} - P_{h'k}^{(n+1)} = \sum_j (p_{hj} - p_{h'j}) P_{jk}^{(n)}.$$

Soient j', celles des valeurs $j = 1, 2, \ldots$, pour lesquelles

$$u_j \equiv p_{hj} - p_{h'j} \geqq 0,$$

j'' les autres On peut écrire

$$P_{hk}^{(n+1)} - P_{h'k}^{(n+1)} = \sum_{j'} |u_{j'}| P_{j'k}^{(n)} - \sum_{j''} |u_{j''}| P_{j''k}^{(n)}.$$

---

(1) Dans le cas du battage des cartes [ou plus généralement lorsque la condition (T') traitée plus loin, p 38, est vérifiée], les limites P_k étant égales à $\frac{1}{r}$ sont nécessairement positives. C'est ce qui explique pourquoi les auteurs qui se sont occupés plus spécialement du battage des cartes, n'ont pas eu à distinguer le cas positivement régulier.

Or

$$\sum_{j'} |u_{j'}| - \sum_{j''} |u_{j''}| = \sum_{j} u_j = \sum_{j} p_{hj} - \sum_{j} p_{h'j} = 1 - 1 = 0.$$

On peut donc poser

$$\theta = \sum_{j'} |u_{j'}| = \sum_{j''} |u_{j''}|,$$

et l'on a

$$(2) \qquad 0 \leqq \theta = \sum_{j'} p_{hj'} - \sum_{j'} p_{h'j'} \leqq \sum_{j} p_{hj} = 1.$$

Considérons d'abord le cas où $\theta > 0$. On pourra écrire, avec Markoff,

$$\mathrm{P}_{hk}^{(n+1)} - \mathrm{P}_{h'k}^{(n+1)} = \theta \left\{ \frac{\sum_{j'} |u_{j'}| \mathrm{P}_{j'k}^{(n)}}{\sum_{j'} |u_{j'}|} - \frac{\sum_{j''} |u_{j''}| \mathrm{P}_{j''k}^{(n)}}{\sum_{j''} |u_{j''}|} \right\} \leqq \theta (\mathrm{P}_{k}^{(n)} - p_{k}^{(n)})$$

Car les deux fractions sont des moyennes pondérées de certains des $\mathrm{P}_{jk}^{(n)}$ et par suite sont comprises entre $\mathrm{P}_{k}^{(n)}$ et $p_{k}^{(n)}$.

Lorsque $\theta = 0$, $p_{hj} - p_{h'j}$ doit être nul pour chaque valeur de j, par suite $\mathrm{P}_{hj}^{(n+1)} - \mathrm{P}_{h'j}^{(n+1)}$ doit aussi être nul et l'on a encore

$$\mathrm{P}_{hk}^{(n+1)} - \mathrm{P}_{h'k}^{(n+1)} \leqq \theta (\mathrm{P}_{k}^{(n)} - p_{k}^{(n)})$$

avec $0 \leqq \theta \leqq 1$, θ dépendant uniquement de h et h'. Ce résultat, qu'on aurait pu déduire directement de (I_1), n'apporte, jusqu'à présent, rien de nouveau. Mais l'origine de θ va nous permettre de le remplacer, dans une hypothèse convenable, par un nombre indépendant de h et h' et effectivement inférieur à l'unité. On a, en effet, en utilisant plus complètement l'inégalité (2)

$$\theta \leqq 1 - \sum_{j'} p_{h'j'} \leqq 1 - \sum_{j'} p_{j'}^{(1)},$$

et de même

$$\theta \leqq 1 - \sum_{j''} p_{hj''} \leqq 1 - \sum_{j''} p_{j''}^{(1)}.$$

Faisons intervenir maintenant la condition nécessaire obtenue plus

haut, ou plutôt, pour commencer, une condition un peu plus stricte, à savoir que $\nu = 1$, c'est-à-dire qu'il existe une valeur l du second indice telle que la ligne des p_{il} dans D soit positive et soit ε le plus petit des éléments de cette ligne. Le nombre l est, soit un des j' et alors

$$\sum_{j'} p_{j'}^{(1)} \geqq p_l^{(1)} = \varepsilon$$

soit un des j'' et alors

$$\sum_{j''} p_{j''}^{(1)} \geqq p_l^{(1)} = \varepsilon$$

Dans les deux cas, on a

$$0 \leqq \theta \leqq 1 - \varepsilon < 1,$$

où ε est un nombre choisi indépendamment de h et h'. Et par suite

$$P_{hk}^{(n+1)} - P_{h'k}^{(n+1)} \leqq (1 - \varepsilon)[P_k^{(n)} - p_k^{(n)}]$$

Le second membre étant indépendant de h et de h', on a donc enfin

$$[P_k^{(n+1)} - p_k^{(n+1)}] \leqq (1 - \varepsilon)[P_k^{(n)} - p_k^{(n)}]$$

et puisque ε est indépendant de n

$$(3) \qquad [P_k^{(n)} - p_k^{(n)}] \leqq (1 - \varepsilon)^{n-1}[P_k^{(1)} - p_k^{(1)}]$$

D'ailleurs P_k étant, comme $P_{hk}^{(n)}$, compris entre $P_k^{(n)}$ et $p_k^{(n)}$, on a

$$|P_k - P_{hk}^{(n)}| \leqq P_k^{(n)} - p_k^{(n)},$$

d'où, finalement,

$$(4) \qquad |P_{hk}^{(n)} - P_k| \leqq (1 - \varepsilon)^{n-1}[P_k^{(1)} - p_k^{(1)}]$$

Remarque. — D'après l'inégalité (4), $|P_{hk}^{(n)} - P_k|$ converge vers zéro au moins aussi vite que les termes d'une progression géométrique. On a vu qu'on pouvait prendre pour raison q de cette progression la valeur $1 - \varepsilon = 1 - p_l^{(1)}$. Mais on peut préciser ce résultat en diminuant q. On avait, en effet,

$$0 = \sum_{j'} p_{hj'} - \sum_{j'} p_{h'j'} = 1 - \sum_{j''} p_{hj''} - \sum_{j'} p_{h'j'} \leqq 1 - \sum_{j} p_j^{(1)}$$

on peut donc prendre $q = 1 - \sum_j p_j^{(1)}$. Comme $\sum_j p_j^{(1)} \geqq p_l^{(1)} = \varepsilon$, on voit qu'on

aura obtenu une raison plus petite que $1-\varepsilon$, s'il existe plusieurs lignes $l, l', \ldots$, où tous les termes sont positifs. Et l'on n'aura jamais obtenu une raison $> 1-\varepsilon$.

Si tous les p_{lk} sont $\geqq \omega > 0$, on peut prendre pour q la valeur $1-2\omega$ d'après M. Paul Lévy [1, p. 49], ou même $1-r\omega$ d'après M. Hadamard [4, note (²), p. 630-631]. Comme alors $1-\sum_l p_l^{(1)} \leqq 1-r\omega \leqq 1-2\omega$, notre résultat est toujours au moins aussi précis Il est même, en général, plus précis C'est, par exemple, ce qui a lieu quand le tableau des p_{lk} est

$$\begin{Vmatrix} \frac{5}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{1}{7} & \frac{2}{7} & \frac{3}{7} \end{Vmatrix}$$

cas où les valeurs à prendre pour q dans les trois estimations sont, dans l'ordre $\frac{5}{7}, \frac{4}{7}, \frac{3}{7}$.

Considérons maintenant le cas le plus général où la condition nécessaire obtenue page 27 se trouve vérifiée, sans supposer $\nu = 1$. S'il existe une valeur ν de n telle que, dans le déterminant $D^{(\nu)}$, l'une des lignes (correspondant à un même second indice l des $P_{il}^{(\nu)}$) soit positive, alors on pourra appliquer le résultat précédent aux suites $P_k^{(\nu)}, P_k^{(2\nu)}, \ldots, P_k^{(u\nu)}, \ldots, p_k^{(\nu)}, p_k^{(2\nu)}, \ldots, p_k^{(u\nu)}, \ldots$ Car on peut prendre pour point de départ la relation récurrente

$$P_{hk}^{[(u+1)\nu]} = \sum_j P_{hj}^{(\nu)} P_{jk}^{(u\nu)},$$

tout à fait analogue à celle sur laquelle est basé le raisonnement précédent. Dès lors, ces deux suites, qui tendent encore vers P_k et p_k, ont aussi la même limite $P_k = p_k$.

De plus, la convergence de $|P_{hk}^{(n)} - P_k|$ vers o est « exponentielle », c'est-à-dire au moins aussi rapide que celle des termes d'une progression géométrique convergente. Cela résulte de l'inégalité (4) dans le premier cas considéré, celui où $\nu = 1$. Dans le cas plus général considéré ensuite, si ε' est le plus petit des $P_{il}^{(\nu)}$, on a d'abord, en prenant, $\alpha_\nu = 1 - \varepsilon'$, ou mieux $\alpha_\nu = 1 - \sum_j p_j^{(\nu)}$,

$$P_k^{(u\nu)} - p_k^{(u\nu)} \leqq (\alpha_\nu)^{u-1} [P_k^{(\nu)} - p_k^{(\nu)}] = (\alpha_\nu)^{u-1} \Delta.$$

Or, si n est un entier quelconque $> \nu$, il existe un entier $u > 0$ tel que

$$u\nu \leqq n < (u+1)\nu,$$

d'où

$$|P_{hk}^{(n)} - P_k| \leqq P_k^{(n)} - p_k^{(n)} \leqq P_k^{(u\nu)} - p_k^{(u\nu)} \leqq (\alpha_\nu)^{u-1} A < (q_\nu)^{n-1} \left[\frac{A q_\nu}{(\alpha_\nu)^2}\right],$$

en posant $q_\nu = \sqrt[\nu]{\alpha_\nu}$.

Donc $|P_{hk}^{(n)} - P_k|$ converge vers zéro au moins aussi rapidement que les termes d'une progression géométrique convergente de raison $q_\nu < 1$.

Il en résulte, en particulier, que : *dans le cas régulier, la série*

$$s_{hk} = \sum_{n=1}^{n=+\infty} [P_{hk}^{(n)} - P_k]$$

est absolument convergente et majorée par une progression géométrique convergente.

Dans le raisonnement de Markoff où il existe un déterminant $D^{(n_0)}$ dont tous les éléments sont positifs, non seulement les P_k existent, mais puisqu'ils sont tous au moins égaux aux $p_k^{(n_0)}$ correspondants, ils sont tous positifs.

Ainsi :

I. **Cas régulier.** — Pour que les $P_{hk}^{(n)}$ tendent respectivement, quand n croît indéfiniment, vers des limites P_k indépendantes du premier indice h, il faut et il suffit que, parmi les tableaux $D^{(n)}$ des $P_{hk}^{(n)}$, il en existe au moins un $D^{(\nu)}$ dont *une ligne*, au moins, ait ses éléments $P_{1l}^{(\nu)}, P_{2l}^{(\nu)}, \ldots, P_{rl}^{(\nu)}$ tous positifs. Et alors, parmi les P_k, P_l au moins n'est pas nul.

II. **Cas positivement régulier.** — Pour qu'en outre toutes les limites P_k soient positives, il faut et il suffit que l'un au moins des $D^{(n)}$ ait tous ses éléments positifs.

On peut encore décrire ces deux conditions de la manière plus suggestive suivante.

Pour qu'on soit dans le cas positivement régulier, il faut et il

suffit qu'il existe un nombre n_0 d'opérations assez grand pour que le passage en n_0 opérations d'un état E_h quelconque à un état E_k quelconque soit possible (ou plus précisément de probabilité positive).

Pour qu'on soit dans le cas régulier, il faut et il suffit que, parmi les différents états du système, il en existe au moins un, E_l, tel qu'en prenant ν assez grand, le passage en ν opérations de l'un quelconque des états du système à E_l soit possible (ou plus précisément de probabilité positive).

Exemples. — Si les p_{hk} sont tous égaux, par suite égaux à $\frac{1}{r}$ en vertu de (T_1), les $P^{(n)}_{hk}$ seront tous aussi égaux à $\frac{1}{r}$; on a ainsi un exemple du cas positivement régulier.

Donnons un exemple d'un cas régulier *non positivement*, soit

$$D \equiv \left\| \begin{matrix} 1 & (1-a) & (1-b) \\ 0 & a & 0 \\ 0 & 0 & b \end{matrix} \right\| \qquad \text{avec} \quad 0 \leq a \leq 1, \; 0 \leq b \leq 1$$

I. Si $a < 1$ et $b < 1$, on a une ligne positive. Donc les $P^{(n)}_{hk}$ doivent avoir des limites indépendantes de h : on est dans le cas régulier.

Mais on va voir que ces limites ne sont pas toutes positives : On n'est pas dans le cas positivement régulier. Formons les équations d'itération : On trouve

$$P^{(n)}_{h2} = a P^{(n-1)}_{h2} \qquad \text{et} \qquad P^{(n)}_{h3} = b P^{(n-1)}_{h3}.$$

Donc

$$P^{(n)}_{h2} = a^{n-1} p_{h2}, \qquad P^{(n)}_{h3} = b^{n-1} p_{h3}.$$

Et alors

$$P^{(n)}_{h1} = 1 - P^{(n)}_{h2} - P^{(n)}_{h3} = 1 - a^{n-1} p_{h2} - b^{n-1} p_{h3}.$$

Comme $a < 1$ et $b < 1$, on a le déterminant limite

$$\begin{vmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}.$$

II. Si $a = 1$, $b < 1$, on a comme déterminant limite

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix}$$

Tous les éléments ont des limites. Mais on n'est plus dans le cas régulier, car ces limites ne sont pas égales dans chaque ligne.

III. Si $a = 1$, $b = 1$, les $P_{hk}^{(n)}$ sont indépendants de n, mais non de h, les déterminants $D^{(n)}$ sont identiques à D, on est encore dans le cas où les $P_{hk}^{(n)}$ ont des limites dépendant de h.

IV. Il en est de même du cas où $a = b = 0$. Enfin, on trouvera page 36, 120 des exemples du cas où les $P_{hk}^{(n)}$ n'ont pas chacun une limite déterminée quand n croît.

Cas de Hostinský. — Lorsque certains des p_{ij} donnés sont nuls, il est intéressant de déterminer des cas aussi généraux que possible où, sans avoir à calculer effectivement les $P_{ij}^{(n)}$, on peut reconnaître que ceux-ci seront tous positifs pour n assez grand. Le probleme du mélange des urnes en fournit, comme nous le verrons, un exemple. On doit à M. Hostinský [6] d'avoir précisé ce qui permet de généraliser cet exemple. Dans ce problème, si l'on forme le déterminant des p_{ij}, on a vu (p. 16) qu'en général la diagonale principale et les deux diagonales adjacentes sont formées de termes positifs, les autres étant nuls. Or, on peut démontrer que s'il en est de même dans le cas plus général de Markoff, le déterminant $D^{(n)}$ des $P_{hk}^{(n)}$ aura tous ses termes positifs à partir d'un certain rang n et l'on sera dans le cas positivement régulier. Nous pouvons, pour le prouver, faire d'abord les remarques suivantes qui, sans être indispensables ici, ont leur intérêt propre.

Tout d'abord, supposons que $p_{hh} > 0$, alors en vertu de (I_1), $P_{hh}^{(2)} \geqq p_{hh} p_{hh} > 0$, $P_{h}^{(3)} \geqq p_{hh} P_{hh}^{(2)} > 0$, ... Ainsi lorsqu'un élément de la diagonale principale de D est positif, il en est de même, quel que soit n, des éléments des $D^{(n)}$ qui occupent la même place.

En outre, *dans le même cas, dans aucun casier de* $D^{(n)}$ *appartenant à la ligne de* p_{hh} *ne peut apparaître un* 0 *quand on passe de* $D^{(n)}$ *à* $D^{(n+1)}$.

Car, d'après (I_1), $P_{hk}^{(n)} \geqq p_{hh} P_{hk}^{(n-1)}$. Si donc $P_{hk}^{(n-1)} > 0$, alors $P_{hk}^{(n)}$ est positif, par suite, $P_{hk}^{(n+1)}$ aussi, etc.

En particulier, *lorsque la diagonale principale de* D *est positive*, non seulement il en est de même de celles de $D^{(2)}$, de $D^{(3)}$, .. , mais, si des termes nuls de $D^{(n)}$ peuvent subsister dans $D^{(n+1)}$, il ne peut apparaître aucun nouveau terme nul.

Si, maintenant, nous nous trouvons dans le cas où la diagonale principale de D et *l'une* des deux diagonales adjacentes ont leurs termes positifs, on voit d'abord qu'il en sera de même dans $D^{(n)}$ quel que soit n. Alors, s'il y a, de plus dans $D^{(n)}$, α diagonales consécutives du même côté de la diagonale principale et ayant leurs termes positifs, il en sera de même pour $D^{(n+1)}$, et de plus on a, par exemple,

$$P_{k+\alpha+1,k}^{(n+1)} \geqq P_{k+\alpha+1,k+1}^{(n)} p_{k+1,k} > 0.$$

Donc $D^{(n+1)}$ aura, non seulement α, mais $\alpha + 1$ diagonales consécutives positives du même côté. En appliquant à D, $D^{(2)}$, $D^{(3)}$, . . , on voit que dans $D^{(n)}$, pour n assez grand, et, plus précisément dans $D^{(r-1)}$, $D^{(r)}$, les termes d'un même côté de la diagonale principale seront positifs.

Dès lors, dans le cas signalé par M. Hostinský, les déterminants $D^{(n)}$ ont bien tous leurs termes positifs à partir d'un certain rang et même à partir du rang $n = r - 1$.

On pourrait se demander s'il ne suffit pas, pour arriver à la même conclusion que la diagonale principale de D soit positive, puisque nous avons déjà reconnu qu'alors aucun o ne peut s'introduire dans l'itération de D Il n'en est rien comme le montre le cas où D est le déterminant $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$ les déterminants $D^{(n)}$ lui sont alors identiques. Le nombre des termes positifs ne décroît pas, mais il ne croît pas non plus. Il en est de même si, l'ordre r du déterminant D étant quelconque, tous ses termes sont nuls sauf ceux de la diagonale principale, nécessairement égaux à 1 en vertu de la relation

$$\sum_j p_{ij} = 1.$$

Dans le cas du mélange des urnes, on a vu que le cas de M. Hostinský se trouve en général réalisé. Il y a toutefois des exceptions que nous avons signalées p. 16. Mais il est facile de s'assurer directement que, même dans ces cas exceptionnels, les $P_{ij}^{(2)}$ sont positifs pour $|i - j| \leqq 2$

et que les $P_{ij}^{(r-1)}$ sont tous $\neq 0$. Ceci cependant suppose, pour le cas exceptionnel où l'on aurait $u=v=B=N$, que la valeur commune de ces nombres de boules est >1. Dans le dernier cas restant, celui où chaque urne ne contiendrait qu'une boule, les deux boules étant de couleurs différentes, il est clair que les déterminants $D^{(n)}$ prendraient successivement les formes

$$D^{(2m+1)} \equiv \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \qquad D^{(2m)} \equiv \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}.$$

Mais c'est un cas qui, en réalité, ne relève pas de la théorie des probabilités, puisqu'il n'y a plus de hasard dans les opérations successives.

En résumé, *le problème du mélange des urnes, quand le hasard y intervient, rentre toujours dans le cas positivement régulier.* Autrement dit, au bout d'un nombre assez grand d'opérations, la probabilité de chaque composition devient à peu près indépendante de la composition initiale.

Généralisation. — La condition de M. Hostinský suffit pour qu'on soit dans le cas positivement régulier, elle n'est pourtant pas nécessaire. On en verra un exemple, page 99. Mais dans ce même cas positivement régulier, comme $D^{(n)}$ a ses termes tous positifs pour n assez grand, on est assuré que, pour n assez grand, $D^{(n)}$ aura sa diagonale principale positive ainsi que les deux diagonales adjacentes : nous dirons qu'alors la *condition de Hostinský généralisée* est vérifiée. On vient de voir *qu'elle est nécessaire pour qu'on soit dans le cas positivement régulier. Elle est aussi suffisante*, car, si les diagonales en question sont positives pour $D^{(\nu)}$, on pourra raisonner sur les $P_{hk}^{(\nu)}$ et $P_{hk}^{(n\nu)}$ comme sur les p_{hk} et les $P_{hk}^{(n)}$. Autrement dit $D^{[(r-1)\nu]}$ aura tous ses termes positifs. Alors on est dans le cas positivement régulier. On n'aura donc pas, en général, à calculer un nombre aussi grand de déterminants $D^{(n)}$, si l'on se contente de vérifier que la condition de M. Hostinský généralisée est satisfaite, que si l'on applique la condition précédemment formulée page 31.

Remarque I. — La condition trouvée page 31 pour caractériser le cas régulier, exige en principe, pour être utilisée, la possibilité de connaître les valeurs des probabilités itérées. Il est clair que c'est là un

problème aussi difficile, en général, et même plus difficile à résoudre que celui de déterminer si les $P_{hk}^{(n)}$ ont des limites P_k. Dans quelques cas particuliers, par exemple, si une ligne des p_{hk} est positive, ou quand on se trouve dans le cas de M. Hostinský, on pourra utiliser la condition au moyen d'une inspection directe des p_{hk}. Dans le cas général, il reste à trouver une condition de régularité qui puisse être appliquée sans avoir à procéder à l'itération des p_{hk}. C'est ce que nous ferons plus loin de plusieurs façons (p. 112 et 197). Mais, comme les raisonnements à cet effet sont assez longs, nous commencerons d'abord par montrer que leurs résultats en valent la peine, en développant les propriétés intéressantes qu'on peut établir dans l'hypothese où l'on se trouve dans le cas régulier.

Remarque II. — La condition de M. Hostinský est une condition suffisante pour le cas régulier et où le fait que toute la diagonale principale est $\neq 0$ joue un rôle essentiel. Il est donc remarquable de pouvoir précisément négliger la valeur d'un terme de la diagonale principale dans l'application du critère d'après lequel il suffit aussi qu'une ligne du tableau des p_{ik} soit toute différente de zéro. Au moins, cette observation est-elle exacte quand on fait abstraction d'un cas simple facile à traiter sans une itération illimitée. Supposons, par exemple,

$$p_{12},\ p_{32},\ \ldots,\ p_{r2}\ \text{tous} \neq 0 \qquad \text{et} \qquad p_{22} = 0.$$

Alors,

$$P_{22}^{(n)} = \sum_{j \neq 2} P_{2j}^{(n-1)} p_{j2} \geqq \varepsilon[1 - P_{22}^{(n-1)}],$$

où ε est le plus petit des nombres positifs p_{j2} pour $j \neq 2$; et

$$P_{k2}^{(n)} = \sum_j p_{kj} P_{j2}^{(n-1)} \geqq p_{k2} P_{22}^{(n-1)} \geqq \varepsilon P_{22}^{(n-1)} \qquad \text{si} \quad k \neq 2$$

Si donc $P_{22}^{(n-1)}$ est $\neq 0$ et $\neq 1$, la deuxième ligne de $D^{(n)}$ sera positive et l'on sera dans le cas régulier. Reste donc à traiter le cas, où pour chaque valeur de n, $P_{22}^{(n)}$ est égal à 0 ou 1.

Or, on a

$$P_{22}^{(2)} \geqq \varepsilon(1 - p_{22}) = \varepsilon > 0.$$

Donc si $P_{22}^{(2)} \neq 1$, on est dans le cas régulier.

D'ailleurs, le cas où $P_{22}^{(2)} = 1$, est bien un cas d'exception. Car dans ce cas, on aurait en vertu de (T) : $P_{2h}^{(2)} = 0$ pour $h \neq 2$, d'où

$$P_{22}^{(2m+1)} = \sum_j P_{2j}^{(2)} P_{j2}^{(2m-1)} = P_{22}^{(2m-1)},$$

$$P_{22}^{(2m+2)} = \sum_j P_{2j}^{(2)} P_{j2}^{(2m)} \quad = P_{22}^{(2m)}.$$

Il suffit donc que $P_{22}^{(2)} = 1$, pour que les $P_{22}^{(n)}$ soient alternativement égaux à zéro et un, et par suite n'aient pas de limite : on cesse d'être dans le cas régulier. Ainsi, dans le cas où une ligne de D est positive, sauf peut-être le terme de cette ligne appartenant à la diagonale principale, on est encore dans le cas régulier, à l'exception du cas où dans $D^{(2)}$ le terme correspondant de la diagonale principale est égal a 1

II — Valeurs des probabilités limites dans le cas régulier.

Le premier problème qui se pose après celui de l'existence d'une limite est celui de la détermination de la valeur de cette limite. Si lorsque n croît indéfiniment, chacun des $P_{hk}^{(n)}$ tend vers une limite, qu'on peut désigner par P_{hk}, ces valeurs limites P_{hk} satisferont aux conditions obtenues en passant à la limite dans les relations (I_1) et (T).

$$P_{hk} = \sum_j P_{hj}\, p_{jk}, \tag{5}$$

$$\sum_j P_{hj} = 1 \tag{6}$$

On observera que d'après les formules (1), on a

$$p_k^{(1)} \leqq p_{hk} \leqq p_k^{(1)}.$$

Dans le cas régulier, les P_{hk} se réduisent aux P_k, et l'on a

$$P_k = \sum_j P_j\, p_{jk}, \tag{7}$$

$$\sum_{j=1}^{j=r} P_j = 1. \tag{8}$$

α). **Cas de la limite constante.** — Il est intéressant de rechercher si les limites de $P_{hk}^{(n)}$ qui, dans le cas régulier, sont indépendantes de l'état initial E_h, pourraient être aussi indépendantes de l'état final E_k. Dans ce cas, la dernière relation montre que *la limite commune* P *serait égale* à $\frac{1}{r}$. On sera donc nécessairement dans le cas positivement régulier. Et l'avant-dernière montre que ce cas ne peut se présenter que si les p_{ik} vérifient, en outre de la condition (T_1), la condition (T'_1) qui lui est analogue

$$(T'_1) \qquad \sum_{j=1}^{j=r} p_{jk} = 1$$

Cette condition exprime que si une opération est faite simultanément à partir de chacun des états $E_1, \ldots, E_r$, chacun des états $E_1, \ldots, E_r$ sera nécessairement obtenu (¹), de sorte que l'effet de cette opération collective consistera simplement en une permutation de ces états entre eux, la permutation identique étant d'ailleurs admise.

Par exemple, dans un coup de battage des cartes, les nouveaux rangs des cartes seront les rangs possibles 1, 2, ..., r, mais peut-être dans un ordre nouveau.

Par contre, dans le problème du mélange des urnes, nous avons vérifié (p. 16) que la probabilité pour qu'on obtienne une composition déterminée C_i n'est pas égale à l'unité.

Avant d'aller plus loin, observons que de la seule condition (T'_1), on peut tirer deux conséquences. D'abord, si (T'_1) est vérifiée, on aura

$$\sum_h P_{hk}^{(n)} = \sum_h \left[\sum_j P_{hj}^{(n-1)} p_{jk}\right] = \sum_j \left[\sum_h P_{hj}^{(n-1)}\right] p_{jk},$$

d'où, en appelant $A_k^{(n)}$ le premier membre,

$$A_k^{(n)} = \sum_j A_j^{(n-1)} p_{jk},$$

et comme

$$A_k^{(1)} = \sum_h p_{hk} = 1,$$

(¹) Ou plus exactement, il y aura, quel que soit le rang k, une probabilité égale à 1 que l'état E_k apparaisse dans l'une au moins de ces opérations simultanées.

on aura

$$A_k^{(2)} = \sum_j p_{jk} = 1,$$

et, en général $A_k^{(n)} = 1$ ou

$$(\mathrm{T}') \qquad \sum_h P_{hk}^{(n)} = 1$$

D'autre part, on déduit de cette égalité

$$r\,p_k^{(n)} \leqq 1 \leqq r\,P_k^{(n)}.$$

D'où, en passant à la limite

$$p_k \leqq \frac{1}{r} \leqq P_k$$

On a vu que si $P_{hk}^{(n)}$ tend vers une limite P indépendante de h et de k, on est nécessairement dans le cas régulier et l'on a (T'_1). Réciproquement, si l'on a la condition (T'_1) dans le cas régulier, alors $\frac{1}{r}$ est compris entre p_k et P_k et $p_k = P_k$. Donc $p_k = P_k = \frac{1}{r}$.

On peut donc dire : pour que les $P_{hk}^{(n)}$ tendent vers une limite indépendante de k et de h, il faut et il suffit que la condition (T'_1) soit vérifiée et qu'il existe au moins un tableau $D^{(\nu)}$ dont au moins une ligne soit positive.

D'ailleurs, puisqu'on permute (T'_1) et (T_1) quand on permute les lignes et les colonnes de D et par suite de $D^{(\nu)}$, on voit qu'on peut remplacer dans l'énoncé précédent une ligne par une colonne.

Finalement : *la condition nécessaire et suffisante pour que les* $P_{hk}^{(n)}$ *tendent vers une limite constante est que la condition* (T'_1) *soit vérifiée et qu'il existe un nombre* ν *d'opérations assez grand pour qu'au moins une ligne ou une colonne de* $D^{(\nu)}$ *soit positive.* Cette seconde condition s'exprime aussi en disant qu'il existe au moins un état E_l tel qu'il soit possible (avec probabilité positive) de passer en ν opérations de E_l à l'un quelconque des états possibles.

Nous appellerons *cas le plus régulier le cas actuel :* celui où les $P_{hk}^{(n)}$ convergent vers une limite indépendante de l'état initial E_h et de l'état final E_k.

Remarque I. — On voit que si la condition (T'_1) est vérifiée, on

ne peut être dans le cas régulier sans être dans le cas positivement régulier. Par suite, dans ce cas, $D^{(n)}$ a pour n assez grand, tous ses termes positifs. Mais on a obtenu plus haut une condition nécessaire et suffisante dont il sera, en général, plus simple de vérifier qu'elle est réalisée.

Remarque II. — Il est intéressant de noter un cas où la réalisation de la condition (T) entraîne celle de la condition (T'), et d'ailleurs réciproquement, c'est celui où p_{hk} est une fonction *symétrique* de h et de k, c'est-à-dire où

$$p_{hk} = p_{kh} \qquad \text{pour} \quad h, k = 1, 2, \ldots, r$$

Application au battage des cartes. — D'après ce qui précède, si les opérations de battage sont telles qu'elles ne s'opposent pas au passage d'une carte d'un rang quelconque à un rang quelconque après une seule opération, alors les probabilités de passage d'un rang h à un rang k au bout de n opérations tendent à s'uniformiser quand le nombre de ces opérations croît. Elles tendent toutes vers l'inverse $\frac{1}{r}$ du nombre des cartes.

C'est, en général, ce qui se passera Toutefois, le joueur peut avoir des habitudes qui ne satisfont pas à cette condition (que les p_{hk} sont tous $\neq 0$). Par exemple, quand on « coupe », la première carte va nécessairement quitter son rang si l'on se borne à une seule opération : p_{11} sera nul.

Nous aurons donc un cas plus général qui pourra effectivement se présenter dans la pratique et où le résultat précédent subsiste, c'est celui où certains des p_{hk} pouvant être nuls, les opérations de battage sont cependant telles qu'on puisse passer d'un rang h quelconque à un rang k quelconque au bout d'un nombre ν suffisant d'opérations, ν étant indépendant de h et de k. Il peut en être ainsi même pour des procédés de battage qui semblent très imparfaits; par exemple, si l'on se bornait à chaque opération à choisir au hasard deux cartes contiguës et à les permuter. On serait alors dans le cas de M. Hostinský, où les p_{hh}, $p_{h,h+1}$, $p_{h,h-1}$ sont les seuls p_{hk} qui sont $\neq 0$. Et l'on a vu qu'alors tous les $P_{hk}^{(r-1)}$ sont positifs. Ainsi avec ce procédé de battage, on finirait encore par uniformiser les probabilités $P_{hk}^{(n)}$. Mais,

bien entendu, leurs différences avec $\frac{1}{r}$ seraient lentes à se réduire. Il faudrait en effet $r - 1$ (en pratique 31) opérations pour arriver seulement à ce résultat qu'elles soient toutes positives.

Mais on peut être aussi, avec d'autres habitudes, dans le cas non régulier. Supposons, par exemple, que le joueur sépare le paquet de cartes en deux paquets égaux, qu'il opère un coup de battage sur chaque paquet séparément, puisqu'il les accole en laissant à droite le paquet de droite. En appelant k' les rangs dans le premier paquet, k'' dans le second, les $p_{k'k''}$ et $p_{k''k'}$ seront nuls. En opérant à nouveau n fois de cette façon, on voit que les $P^{(n)}_{k'k''}$, $P^{(n)}_{k''k'}$ resteront nuls. On sera dans le même cas que si chaque paquet était battu séparément. Si les opérations de battage sont telles que les $p_{k'k'_1}$ et les $p_{k''k''_1}$ soient tous positifs, alors on sera, pour chaque paquet, dans le cas même examiné page 38, et l'on aura

$$\lim_{n \to \infty} \left\{ \begin{array}{l} P^{(n)}_{k'k'} = \frac{2}{r}, \\ P^{(n)}_{k''k''} = \frac{2}{r}, \\ P^{(n)}_{k'k''} = 0, \\ P^{(n)}_{h''k'} = 0 \end{array} \right.$$

Les $P^{(n)}_{hk}$ ont bien dans ce cas des limites, mais ces limites ne sont ni indépendantes du premier indice, ni égales, ni même toutes positives.

C'est là un cas où il serait clair pour tous les spectateurs que le battage est imparfait. Il y a d'autres cas où ce serait moins évident. Un joueur consciencieux pourrait procéder de la façon suivante : séparer les cartes en deux paquets, introduire le paquet de droite *dans* le paquet de gauche de façon que les unes viennent s'imbriquer dans les autres. Il peut sembler qu'un nombre suffisant de ces opérations devrait aboutir à un bon mélange des cartes. On n'aura peut-être pas, en effet, observé que, dans toutes ces opérations, la première carte, la carte la plus à gauche, reste inchangée. De sorte que les $P^{(n)}_{12}$, $P^{(n)}_{13}, \ldots, P^{(n)}_{1r}$, $P^{(n)}_{21}$, $P^{(n)}_{31}, \ldots, P^{(n)}_{r1}$ restent constamment nuls et $P^{(n)}_{11}$ constamment égal à l'unité. $P^{(n)}_{11}$, $P^{(n)}_{21}$ ayant alors pour limites respectives 1 et 0, on ne peut être dans le cas régulier. C'est, à vrai dire, un cas analogue au précédent, mais un cas où l'un des

deux paquets ne contient qu'une carte et où la séparation en deux paquets a pu se faire de façon inconsciente.

Autre position du problème — Il y a d'ailleurs une autre façon de poser le problème du battage des cartes. Au lieu de porter son attention sur chacune des cartes séparément, on peut, avec Poincaré, étudier les changements dans l'ordre de *l'ensemble* des cartes. Pour un paquet de N cartes, il y a un nombre de dispositions différentes de ces cartes égal au nombre N ! des permutations U de leurs rangs. En posant $r = N\,!$ et en numérotant $U_1, U_2, \ldots, U_r$ ces différentes permutations, il s'agira d'étudier le comportement quand n croît de la probabilité $\varpi_{hk}^{(n)}$ que le paquet passe de la permutation U_h à la permutation U_k au bout de n opérations. On a encore $\sum_{k=1}^{k=r} \varpi_{hk}^{(n)} = 1$, car une opération transforme nécessairement la permutation U_h en l'une des permutations $U_1, \ldots, U_r$. La discussion du problème général de Markoff nous conduit dès maintenant, quand on l'applique au battage des cartes, au premier résultat suivant.

Si les *habitudes* du joueur sont telles qu'elles rendent possible le passage en un même nombre fini convenable d'opérations, de l'une quelconque des permutations à l'une des autres, par exemple, de U_1, alors les $\varpi_{hk}^{(n)}$ tendront vers des limites ϖ_k indépendantes de la permutation initiale U_h, et l'on aura $\sum_k \varpi_k = 1$.

Mais nous allons pouvoir préciser. Quand nous avons considéré chaque carte isolément, alors, en désignant, par $P_{hk}^{(n)}$, la probabilité de passer du rang h au rang k, nous avons supposé implicitement que cette probabilité était indépendante de la *nature* de la carte. Considérons l'ensemble des cartes; la substitution qui fait passer de la permutation U_h à la permutation U_k est aussi celle qui fait passer de la permutation $1, 2, \ldots, r$ représentée par U_1 à une certaine permutation U_j. On peut représenter cette correspondance par la notation $U_j = U_h U_k$. Si nous restons dans la même hypothèse qui vient d'être formulée, la probabilité de passer de U_h à U_k doit être la même que celle de passer de U_1 à U_j (dans le même nombre d'opérations). On pourra poser $P_j^{(n)} = \varpi_{hk}^{(n)}$ et l'on aura évidemment

$\sum_j P_j^{(n)} = 1$. Il est clair que si, dans l'identité $U_i = U_h U_k$, on laisse fixe h ou k et l'on donne à l'autre toutes les valeurs de 1 à r, j prendra aussi des valeurs distinctes qui iront de 1 à r. Dès lors, on a

$$\sum_h \varpi_{hk}^{(n)} = \sum_k \varpi_{kh}^{(n)} = 1.$$

Il en résulte, comme on l'a vu plus haut, que ϖ_h sera aussi indépendant de k et par suite égal à $\frac{1}{r} = \frac{1}{N!}$ (et non plus à l'inverse du nombre N des cartes). Ainsi, dans l'hypothèse où les habitudes du joueur sont indépendantes de la nature des cartes qui figuraient initialement aux rangs 1, 2, .., N et où les probabilités de passer d'au moins une permutation telle que U_1 à une autre quelconque sont toutes positives (au moins après un même nombre fini d'opérations), *ces probabilités tendent vers un même nombre, égal à* $\frac{1}{N!}$ *quand le nombre des opérations croît indéfiniment.*

M. Hadamard a traité aussi le cas singulier sur lequel nous reviendrons plus loin (p. 199).

β). **Valeurs non nécessairement égales des probabilités limites P_h.** — Retournons maintenant au cas général du problème de Markoff : celui où la condition (T′) n'est pas nécessairement vérifiée. Sans faire d'hypothèse sur le nombre et la situation des termes non nuls de D ou de $D^{(2)}$, ... ou de $D^{(n)}$, .., si $P_{hk}^{(n)}$ tend vers une limite déterminée P_{hk} lorsque n croît indéfiniment, alors les relations (I_1) et (T_1) montrent que ces limites satisfont aux équations

$$P_{hk} = \sum_{j=1}^{j=r} P_{hj}\, p_{jk} \qquad (h, k = 1, 2, \ldots, r),$$

$$\sum_{j=1}^{j=r} P_{hj} = 1 \qquad (h = 1, 2, \ldots, r).$$

Pour chaque valeur de h, P_{h1}, ..., P_{hr} sont donc un système de

solutions des équations

$$(\mathcal{E}) \qquad \begin{cases} x_k = \sum_j x_j p_{jk} & (k = 1, \ldots, r), \\ \sum_j x_j = 1 \end{cases}$$

Il y a là $r+1$ équations à r inconnues. Mais ces équations ne sont pas indépendantes car, en ajoutant les r premières, on trouve l'identité $\sum_k x_h = \sum_j x_j$, en tenant compte de (T). Il reste donc un système de r équations à r inconnues.

En général, ce système n'a qu'un système de solutions, alors si les $P_{hk}^{(n)}$ convergent, leurs limites $P_{h1}, \ldots, P_{hr}$ seront respectivement égales à ces solutions et, comme ces équations en x ne dépendent pas de h, $P_{h1}, \ldots, P_{hr}$ seront nécessairement indépendantes de h. Si les limites P_{hk} existent, elles sont, dans ce cas, bien déterminées, et l'on est nécessairement dans le cas régulier.

Cependant, il peut arriver que le système $(\mathcal{E})$ ait plus d'un système de solutions. Alors le raisonnement fait précédemment tombe en défaut; tout ce qu'on peut dire pour le moment (voir pour plus de détails, p. 111) c'est que si les $P_{hk}^{(n)}$ ont des limites déterminées, alors, pour chaque valeur de h, ce sont des solutions du système $(\mathcal{E})$, mais la réciproque n'est pas vraie.

Par exemple, dans le cas considéré plus haut où les termes de D sont nuls, sauf la diagonale principale, alors les relations (I_1) deviennent

$$P_{hk}^{(n)} = p_{hh} P_{hk}^{(n-1)} = P_{hk}^{(n-1)},$$

d'où

$$P_{hk}^{(n)} = p_{hk} = \begin{cases} 1 & \text{si } h = k, \\ 0 & \text{si } h \neq k \end{cases}$$

Par suite, dans ce cas $\lim_{n \to \infty} P_{hk}^{(n)} = p_{hk}$. la limite existe, mais n'est pas indépendante de h. Pour chaque valeur de h, le système de limites $P_{h}, \ldots, P_{hr}$ est l'un des systèmes de solutions du système $(\mathcal{E})$ qui est ici

$$x_k = x_k,$$
$$\Sigma x_k = 1,$$

et qui a par conséquent beaucoup d'autres systèmes de solutions que les r systèmes de solutions $x_1 = p_{h1}, \ldots, x_r = p_{hr}$, $(h = 1, \ldots, r)$

Même lorsque le système ($\mathcal{E}$) n'a qu'un système de solutions, ce système ne fournit pas nécessairement les limites des $P_{hk}^{(n)}$. Il faut pour cela que ces limites existent.

Prenons comme exemple le cas où le déterminant D est $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$. Alors $D^{(2)}$ sera $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$, $D^{(3)}$ sera $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$, ..., en général $D^{(2n+1)} \equiv D$; $D^{(2n)} = D^{(2)}$. Par suite $P_{hk}^{(n)}$ prend alternativement quand n croît les valeurs 0 et 1. $P_{hk}^{(n)}$ n'a pas de limite déterminée quand n croît. Pourtant, dans ce même cas, les équations ($\mathcal{E}$) ont une solution unique. Elles deviennent en effet

$$x_1 = x_2, \qquad x_2 = x_1, \qquad x_1 + x_2 = 1 \qquad \text{d'où} \quad x_1 = x_2 = \frac{1}{2}.$$

Nous verrons, page 111, que si le système ($\mathcal{E}$) n'a qu'un système de solutions, on peut donner une interprétation simple de ces solutions même quand il n'y a pas de limite des $P_{hk}^{(n)}$ au sens ordinaire. On indiquera aussi (p. 114) comment on peut résoudre ces équations de façon simple.

Nous allons cependant montrer que, *dans le cas régulier, le système des équations* ($\mathcal{E}$) *n'a qu'un seul système de solutions*. En effet, plaçons-nous dans le cas régulier, et soit $X_1, \ldots, X_r$ un système de solutions du système ($\mathcal{E}$) (Nous savons qu'il y a au moins le système de solutions $P_1, \ldots, P_r$.) On aura

$$X_k = \sum_j X_j p_{jk} = \sum_j \left[\sum_i X_i p_{ij} \right] p_{jk} = \sum_i X_i \sum_j p_{ij} p_{jk} = \sum X_i P_{ik}^{(2)}.$$

De même, en itérant n fois

$$X_k = \sum_i X_i P_{ik}^{(n)},$$

et en faisant croître n indéfiniment

$$X_k = \sum_i X_i P_k = P_k \sum_i X_i = P_k$$

Ainsi $X_1 = P_1, \ldots, X_r = P_r$.

Cette remarque, très importante en soi, va nous permettre également de déterminer les quantités s_{ik} que nous avons définies précédemment (p. 31) et qui interviendront à plusieurs reprises. Par défi-

nition dans le cas régulier :

$$s_{ik} = \sum_{t=1}^{t=\infty} [P_{ik}^{(t)} - P_k].$$

Or,

$$P_{ik}^{(t+1)} = \sum_j P_{ij}^{(t)} p_{jk},$$

d'où, quand t croît,

$$P_k = \sum_j P_j p_{jk},$$

et

$$P_{ik}^{(t+1)} - P_k = \sum_j (P_{ij}^{(t)} - P_j) p_{jk},$$

$$\sum_{t=1}^{t=n} (P_{ik}^{(t+1)} - P_k) = \sum_j \left[\sum_{t=1}^{t=n} (P_{ij}^{(t)} - P_j) \right] p_{jk}$$

D'où, quand n croît,

$$(8\ bis) \qquad s_{ik} - (p_{ik} - P_k) = \sum_j s_{ij} p_{jk}$$

De même

$$\sum_j \sum_{t=1}^{t=n} (P_{ij}^{(t)} - P_j) = \sum_{t=1}^{t=n} \left[\sum_j P_{ij}^{(t)} - \sum_j P_j \right] = \sum_t (1 - 1) = 0,$$

et à la limite

$$(9) \qquad \sum_j s_{ij} = 0.$$

Par suite $s_{i1}, \ldots, s_{ir}$ sont solutions du système d'équations linéaires :

$$(\mathcal{E}'_i) \qquad \begin{cases} z_k - (p_{ik} - P_k) = \sum_j z_j p_{jk} & (k = 1, \ldots, r), \\ \sum_j z_j = 0. \end{cases}$$

On voit, de même que pour le système ($\mathcal{E}$), que les r premières équations ne sont pas indépendantes, de sorte que le système ($\mathcal{E}'_i$) se réduit à r équations.

Or, le déterminant des coefficients de ce système est le même que

celui du système ($\mathcal{E}$). Par suite, le système ($\mathcal{E}'_i$) n'a dans le cas régulier qu'un système de solutions, à savoir :

$$z_1 = s_{i1}, \qquad z_r = s_{ir}$$

Lorsqu'on fait intervenir les

$$s_{ij} = \sum_{t=1}^{t=\infty} [P_{ij}^{(t)} - P_j]$$

dans un calcul, il semble, en raison de la définition des s_{ij}, qu'il faudrait, pour les déterminer, connaître tous les $P_{ij}^{(t)}$. Nous voyons maintenant que dans le cas régulier, *on peut déterminer les quantités s_{ij} sans itération*, à partir des données p_{jk} au moyen de ($\mathcal{E}'_i$) [où les P_k qui figurent dans ($\mathcal{E}'_i$) se déterminent eux aussi sans itération au moyen du système linéaire ($\mathcal{E}$)].

En opérant comme plus haut après permutation de P et p, on aurait aussi trouvé :

$$(8\ ter) \qquad s_{ik} - (p_{ik} - P_k) = \sum_j p_{ij} s_{jk}$$

La relation analogue à $\sum_j s_{ij} = 0$ est moins simple.

Partons de l'égalité

$$P_{ik}^{(t+m)} = \sum_j P_{ij}^{(m)} P_{jk}^{(t)}.$$

D'où, quand t croît,

$$P_k = \sum_j P_{ij}^{(m)} P_k.$$

Et en retranchant

$$P_{ik}^{(t+m)} - P_k = \sum_j P_{ij}^{(m)} [P_{jk}^{(t)} - P_k]$$

D'où

$$\sum_{t=1}^{t=m+n} (P_{ik}^{(t)} - P_k) - \sum_{t=1}^{t=m} [P_{ik}^{(t)} - P_k] = \sum_j P_{ij}^{(m)} \left[\sum_{t=1}^{t=n} (P_{jk}^{(t)} - P_k) \right].$$

Quand n croît :

$$s_{ik} - \sum_{t=1}^{t=m} [P_{ik}^{(t)} - P_k] = \sum_j P_{ij}^{(m)} s_{jk}.$$

Et quand m croît

$$0 = s_{ik} - s_{ik} = \sum_j P_j s_{jk}.$$

Ainsi

(9 *bis*)
$$\sum_{j=1}^{j=r} P_j s_{jk} = 0, \qquad (k = 1, \ldots, r)$$

Remarque. — On vient de voir (p. 46), qu'en posant

$$R_{ik}^{(n)} = \sum_{t=1}^{t=n} [P_{ik}^{(t)} - P_k],$$

on a

$$R_{ik}^{(n+1)} - (p_{ik} - P_k) = \sum_j R_{ij}^{(n)} p_{jk} = \sum_j p_{ij} R_{jk}^{(n)},$$

et que $\lim\limits_{n \to \infty} R_{ik}^{(n)} = s_{ik}$.

On peut chercher comment se comporte la quantité $R_{ik}^{(n)} - s_{ik}$. En l'appelant $x_i(n)$, on voit qu'elle vérifie l'équation

$$x_i(n) + s_{ik} - p_{ik} + P_k = \sum_j p_{ij} x_j(n-1) + \sum p_{ij} s_{jk}$$

D'où, en vertu de (8 *ter*).

$$x_i(n) = \sum_j p_{ij} x_j(n-1),$$

avec

$$\lim_{n \to \infty} x_i(n) = 0.$$

Dès lors

$$x_i(n) = \sum_j P_{ij}^{(n-1)} x_j(1),$$

et à la limite,

$$0 = \sum_j P_j x_j(1),$$

d'où

(9 *ter*)
$$x_i(n) = \sum_j [P_{ij}^{(n-1)} - P_j] x_j(1).$$

Donc

$$|x_i(2)| + \ldots + |x_i(n)| \leqq \sum_j |x_j(1)| \sum_{t=2}^{t=n} |P_{ij}^{(t-1)} - P_j|.$$

En posant $R_{ik}^{(n)} = s_{ik} + \varepsilon_{ik}(n)$, on voit que la série $\sum_{n=1}^{n=+\infty} \varepsilon_{ik}(n)$ est absolument convergente, fait qui nous sera utile plus loin. Et sa somme est, en vertu de (9 *ter*),

$$R_{ik}^{(1)} - s_{ik} + \sum_{l=2}^{\infty} \sum_{j} [P_{ij}^{(l-1)} - P_j][R_{jk}^{(1)} - s_{jk}]$$

$$= R_{ik}^{(1)} - s_{ik} + \sum_{j} s_{ij}[R_{jk}^{(1)} - s_{jk}] = p_{ik} - P_k - s_{ik} + \sum_{j} s_{ij}(p_{jk} - P_k - s_{jk})$$

$$= p_{ik} - P_k - s_{ik} + s_{ik} - p_{ik} + P_k - \sum_{j} s_{ij} s_{jk} = -s_{ik}^{(2)},$$

en posant

$$s_{ik}^{(2)} = \sum_{j} s_{ij} s_{jk}$$

Application au mélange des urnes. — $P^{(r-1)}$ ayant, dans ce problème, ses termes positifs, on conclut que les $P_{hk}^{(n)}$ ont aussi des limites, toutes $\neq 0$ et indépendantes de h. Ces limites satisfont aux équations ($\mathcal{E}$) de la page 44, qu'on peut écrire ici

$$x_k = x_{k-1} p_{k-1,k} + x_k p_{k,k} + x_{k+1} p_{k+1,k} \qquad (k = 0, 1, \ldots, r)$$

et

$$\sum_{k=0}^{k=r} x_k = 1.$$

Les premières peuvent s'écrire

$$x_{k+1} = \frac{x_k(1 - p_{kk}) - x_{k-1} p_{k-1,k}}{p_{k+1,k}},$$

en commençant par

$$x_1 = \frac{x_0(1 - p_{00})}{p_{10}} = x_0 \frac{p_{01}}{p_{10}},$$

et finissant par

$$x_r = \frac{p_{r-1,r} x_{r-1}}{1 - p_{rr}}.$$

En se donnant arbitrairement x_0, on déduira de la première x_1, de la seconde x_2, ..., de la dernière x_r. Les formules se simplifient

en tenant compte successivement des relations $\sum_k p_{jk} = 1$. On trouve ainsi

$$x_1 = x_0 \frac{p_{01}}{p_{10}},$$

$$x_2 = \frac{x_0 p_{01} p_{12}}{p_{10} p_{21}},$$

$$\cdots$$

$$x_k = \frac{x_0 p_{01} p_{12} \cdots p_{k-1,k}}{p_{10} p_{21} \cdots p_{k,k-1}},$$

$$\cdots$$

$$x_r = \frac{x_0 p_{01} p_{12} \cdots p_{r-1,r}}{p_{10} p_{21} \cdots p_{r,r-1}}.$$

En les portant dans $\sum_k x_k = 1$, on détermine x_0 par

$$\begin{aligned} x_0 [p_{01} p_{21} \cdots p_{r,r-1} + p_{01} p_{12} p_{32} \cdots p_{r,r-1} + \cdots \\ + p_{01} p_{21} \cdots p_{k-1,k} p_{k+1,k} \cdots p_{r,r-1} + \\ + p_{01} p_{12} \cdots p_{r-1,r}] = p_{10} p_{21} \cdots p_{r,r-1} \end{aligned}$$

Il y a donc un système unique de solutions qui sont nécessairement les valeurs de $P_0, P_1, \ldots, P_r$.

Les formules précédentes fournissant ces valeurs restent valables *dans le cas général où les p_{ih} sont nuls pour* $|i-h|>1$ et $\neq 0$ pour $|i-h| \leqq 1$. Si l'on substitue aux p_{ih} les valeurs obtenues, page 15, dans le cas particulier du mélange des urnes, on trouve, quand $B \leqq v$,

$$x_k = x_0 \frac{C_B^k C_u^k}{C_{v-B+k}^k} = x_0 \frac{C_B^k C_N^{u-k}}{C_N^u}.$$

D'où

$$\frac{x_0}{C_N^u} \{ C_N^u + C_B^1 C_N^{u-1} + \cdots + C_B^k C_N^{u-k} + \cdots + C_B^r C_N^{u-r} \} = 1.$$

Or, r est le plus petit des nombres u et B. Supposons $u \leqq B$; alors $r = u$ et dans ce cas, l'accolade est le coefficient de z^u dans le développement du produit $(1+z)^B (1+z)^N$, c'est-à-dire de $(1+z)^T$ en posant $T = B + N$. L'accolade est donc égale à C_T^u. D'où

$$x_0 = \frac{C_N^u}{C_T^u}.$$

Si l'on avait supposé $B \leqq u$, on aurait écrit

$$x_k = \frac{C_u^k C_v^{B-k}}{C_v^B} x_0,$$

d'où

$$\frac{x_0}{C_v^B}\{C_v^B + C_u^1 C_v^{B-1} + \quad + C_u^r C_v^{B-r}\} = 1,$$

et comme maintenant $r = B$, l'accolade serait égale à C_T^B, d'où

$$x_0 = \frac{C_v^B}{C_T^B}.$$

Or, dans les deux cas, on vérifie facilement que

$$\frac{C_v^B}{C_T^B} = \frac{C_N^u}{C_T^u}.$$

Ainsi, en supposant simplement $B \leqq v$, on obtient dans les deux cas

$$(10)\quad \begin{cases} P_0 = \dfrac{C_N^u}{C_T^u} = \dfrac{C_v^B}{C_T^B}, \\[2ex] P_k = \dfrac{C_B^k C_N^{u-k}}{C_T^u} = \dfrac{C_u^k C_v^{B-k}}{C_T^B}, \\ \dots\dots\dots\dots\dots\dots, \\ P_r = \dfrac{C_B^r C_N^{u-r}}{C_T^u} = \dfrac{C_u^r C_v^{B-r}}{C_T^B} = \begin{cases} \dfrac{C_u^B}{C_T^B} & \text{si } B \leqq u, \\[2ex] \dfrac{C_B^u}{C_T^u} & \text{si } u \leqq B. \end{cases} \end{cases}$$

L'expression de P_k ci-dessus a une signification classique, d'où nous pouvons déduire que : *la probabilité pour qu'au bout de n opérations il y ait k boules blanches dans l'urne* U *tend, quand n croît indéfiniment, vers la probabilité qu'en versant* U *et* V *dans la même urne et en tirant, au hasard, de cette dernière, un nombre de boules égal au nombre initial u de boules de* U, *il y ait parmi celles-ci k boules blanches.*

Composition la plus probable. — Pour quelle valeur de k, P_k a-t-il son maximum ?

On a

$$\frac{P_k}{P_{k-1}} = \frac{C_B^k C_N^{u-k}}{C_B^{k-1} C_N^{u-k+1}} = \frac{(B+1-k)(u+1-k)}{k(N-u+k)}.$$

D'où $P_{k-1} \leqq P_k$ pour $k \leqq \mathcal{E}$, en appelant $\mathcal{E}$ la partie entière de $\frac{(B+1)(u+1)}{T+2}$. En général, ce dernier nombre n'est pas entier, alors P_k croît d'abord quand k croît, atteint son maximum quand $k = \mathcal{E}$, puis décroît. (Quand le même rapport est entier, le résultat sera le même sauf que le maximum sera atteint pour $k = \mathcal{E} - 1$ et $k = \mathcal{E}$.)

Mais quand n croît, les $P_{hk}^{(n)}$ tendent vers les P_k, soit η la plus petite des différences entre le maximum $P_{\mathcal{E}}$ de P_k et les autres valeurs de P_k. Il y a un nombre n_0 indépendant de h et de k tel que pour $n \geqq n_0$ $|P_{hk}^{(n)} - P_k| < \frac{\eta}{3}$. Alors

$$P_{hk}^{(n)} < P_k + \frac{\eta}{3}, \qquad P_{h\mathcal{E}}^{(n)} > P_{\mathcal{E}} - \frac{\eta}{3},$$

et si $k \neq \mathcal{E}$ $\left(\text{et, éventuellement, } k \neq \mathcal{E} - 1 \text{ si } \frac{(B+1)(u+1)}{T+2} \text{ est entier}\right)$ on a

$$P_{\mathcal{E}} \geqq P_k + \eta, \qquad \text{d'où} \quad P_{h\mathcal{E}}^{(n)} > P_{hk}^{(n)} + \frac{\eta}{3}.$$

Dès lors, la composition la plus probable, à partir d'une composition arbitraire C_h est, à partir du rang n_0, une même composition indépendante à la fois de h et de n, c'est la composition $C_{\mathcal{E}}$ (ou éventuellement $C_{\mathcal{E}}$ et $C_{\mathcal{E}-1}$).

La composition la plus probable est ainsi rattachée à la partie entière de $\frac{(B+1)(u+1)}{T+2}$ dont la signification n'est pas immédiate. On remarque cependant que si B et u sont grands, ce nombre est voisin de $\frac{Bu}{T}$ dont nous allons préciser plus loin la signification. Mais nous allons même nous débarrasser de la condition que u et B soient grands.

Dans ce but, posons

$$\mathcal{E} = \frac{Bu}{T} \frac{\left(1+\frac{1}{u}\right)\left(1+\frac{1}{B}\right)}{\left(1+\frac{2}{T}\right)} + \theta,$$

on aura $0 \leqq \theta < 1$. On peut aussi appeler $\mathcal{E}'$ la partie entière de $\frac{Bu}{T}$ et poser $\mathcal{E}' = \frac{Bu}{T} + \theta'$ avec $0 \leqq \theta' < 1$. D'où

$$\mathcal{E} - \mathcal{E}' = \frac{Bu}{T}\left[\frac{\frac{1}{u} + \frac{1}{B} + \frac{1}{uB} - \frac{2}{T}}{1+\frac{2}{T}}\right] + \theta - \theta' = \frac{\frac{Bv}{T} + \frac{Nu}{T} + 1}{T+2} + \theta - \theta'.$$

Or,

$$\frac{Bv}{T}+\frac{Nu}{T}+1 \leqq B+N+1 < T+2$$

Donc

$$\mathcal{E}-\mathcal{E}'=\omega+\theta-\theta' \qquad \text{où} \quad 0<\omega<1,$$

et, par suite,

$$-1<-\theta'<\omega-\theta' \leqq \mathcal{E}-\mathcal{E}' \leqq \omega+\theta<2,$$

$$-1<\mathcal{E}-\mathcal{E}'<2.$$

C'est-à-dire $\mathcal{E}=\mathcal{E}'$ ou $\mathcal{E}=\mathcal{E}'+1$. La composition la plus probable est donc $C_{\mathcal{E}'}$ ou $C_{\mathcal{E}'+1}$. (Dans le cas où il y a deux compositions les plus probables, $\theta=0$ et l'on a $-1<\mathcal{E}-\mathcal{E}'<1$, d'où $\mathcal{E}'=\mathcal{E}$, ces deux compositions sont $C_{\mathcal{E}'}$ ou $C_{\mathcal{E}'+1}$.)

Mais $\frac{Bu}{T}$ est le nombre de boules blanches qu'il devrait y avoir dans U pour que la proportion des blanches y fût la même que dans l'ensemble des deux urnes Ce nombre, en général, n'est pas entier, mais nous voyons qu'il diffère en général de moins d'une unité du nombre de boules blanches de U le plus probable. Plus précisément, nous pouvons dire que, non seulement à la limite, mais *à partir d'un nombre d'opérations suffisamment grand 1° la proportion la plus probable des boules blanches dans chaque urne restera constamment la même, 2° ce sera l'une des deux proportions possibles qui se rapprochent le plus (l'une par excès, l'autre par défaut) de la proportion des boules blanches dans l'ensemble des deux urnes.*

Dans le cas de Laplace, où $u=v=B=N$ et, par suite, $r=u=B$, on aura en particulier

$$(11) \quad P_0=\frac{1}{C_{2u}^u}, \qquad P_1=\frac{(C_u^1)^2}{C_{2u}^u}, \qquad \ldots, \qquad P_k=\frac{(C_u^k)^2}{C_{2u}^u}, \qquad \ldots, \qquad P_u=\frac{1}{C_{2u}^u}.$$

On voit que, dans ce cas, deux compositions C_k, C_{u-k} où la répartition des blanches et des noires est simplement inversée sont à la limite *également probables*. En outre, la variation des probabilités P_k est plus rapide que dans la loi dite binomiale puisque les P_k sont proportionnelles aux *carrés* des coefficients binomiaux Enfin, la répartition la plus probable s'obtient, à partir d'un nombre d'opérations assez grand, pour la proportion $\frac{k}{u}$ la plus voisine de $\frac{1}{2}$. Si, par exemple, $u=2s$, ce serait pour $k=s$. Il faut d'ailleurs observer que, si u est grand, la plus grande P_s des probabilités P_k, ne serait pas très

grande, elle tend même vers zéro avec $\frac{1}{s}$. Car, pour $u = 2s$

$$P_s = \frac{(C_{2s}^s)^2}{C_{4s}^{2s}} = \left[\frac{(2s)!}{s!}\right]^4 \frac{1}{(4s)!},$$

d'où, en appliquant la formule de Stirling, $P_s \backsimeq \sqrt{\frac{2}{\pi s}}$.

Interprétation analytique de l'itération des probabilités au moyen des groupes de transformations. — On peut donner à la relation

$$\text{(I)} \qquad P_{hk}^{(m+n)} = \sum_j P_{hj}^{(m)} P_{jk}^{(n)}$$

une interprétation mathématique intéressante qui se retrouve dans le cas plus général des suites continues d'états et qui est due à M. Hadamard. Considérons en effet une famille $\mathcal{F}$ de transformations linéaires et homogènes $\mathfrak{T}_1$, $\mathfrak{T}_2$ ..., $\mathfrak{T}_m$, ... $\mathfrak{T}_n$, ... de la forme

$$(\mathfrak{T}_n) \qquad z_k = \sum_{j=1}^{j=r} P_{jk}^{(n)} y_j,$$

$$(\mathfrak{T}_m) \qquad y_j = \sum_{h=1}^{h=r} P_{hj}^{(m)} x_h$$

Si l'on effectue sur les y_j une transformation $\mathfrak{T}_n$ de la famille $\mathcal{F}$, les y_j deviennent des z_k et la transformation U des x_h dans les z_k est une transformation linéaire et homogene dont les coefficients sont les seconds membres de (I). L'équation fonctionnelle (I) est donc la condition nécessaire et suffisante pour que : 1° U appartienne à $\mathcal{F}$, c'est-à-dire que $\mathcal{F}$ forme un groupe ; 2° U étant alors du type $\mathfrak{T}_p$, on ait $p = m + n$.

En particulier, si les $p_{ik} = P_{ik}^{(1)}$ satisfont aux conditions de la page 31, si, par suite, on est dans le cas régulier, alors on voit que l'effet cumulé des transformations $\mathfrak{T}_1$ est, pour ainsi dire, de niveler les x_h. De sorte que, si les x_h deviennent des $x_h^{(n)}$ après $\mathfrak{T}_n$, alors, en posant

$$X_h = \lim_{n \to \infty} x_h^{(n)} = \sum_{h=1}^{h=r} P_h x_h,$$

on aura

$$X_1 = X_2 = \quad = X_r = \sum_h P_h x_h$$

Dans le cas particulier où l'on a aussi la condition,

$$(\mathrm{T}'_1) \qquad \sum_h p_{hj} = 1,$$

on aura même

$$X_k = \frac{x_1 + \quad + x_r}{r}.$$

Enfin, on observera que dans le cas général la transformation $\mathcal{T}_n$ n'est autre que celle qui fait passer d'une variable aléatoire X (ayant la valeur x_h pour l'événement E_h) à une variable ayant pour valeur relativement à E_j, la valeur moyenne de X après n épreuves réalisées à partir de l'état E_j.

III. — Probabilités absolues et probabilités inverses.

Loi de probabilité initiale — La transformation $\mathcal{T}_n$ peut être interprétée aussi d'une autre façon quand les x_h sont ≥ 0 et ont une somme égale à l'unité. On peut alors considérer les x_h comme des probabilités et les désigner par ϖ_h. Alors l'expression obtenue dans $(\mathcal{T}_n)$, soit

$$(12) \qquad \varpi_j^{(n)} = \sum_{h=1}^{h=r} P_{hj}^{(n)} \varpi_h,$$

aura la signification suivante. Au lieu de partir d'un état E_k déterminé, on suppose que l'état initial du système est un état aléatoire E_h soumis à la loi de probabilité ϖ_h précisée par les valeurs $\varpi_1, \ldots, \varpi_r$ des ϖ_h; il y a *ensuite* et constamment, une probabilité déterminée p_{hj} de passer en une épreuve de E_h à E_j. De même qu'il y a une probabilité $P_h^{(n)}$ de passer en n épreuves de E_h à E_j, il y aura une probabilité $\varpi_j^{(n)}$ déterminée par la formule (12) d'arriver à l'état E_j en n épreuves, l'état initial étant déterminé par le hasard, soumis à la loi de probabilité ϖ_h.

Par exemple, dans le problème du mélange des urnes, admettons

qu'on ait $u < B$ et $B > 3$, on peut supposer qu'avant de commencer on a tiré d'une troisième urne contenant $B - 3$ boules blanches et N noires, $u - 3$ boules qu'on a placées dans U, qu'on y a ajouté 3 blanches, qu'on a versé enfin les v autres dans V. Ou encore, on pourrait tirer d'une quatrième urne contenant B blanches et N noires, u boules qu'on place dans U et le reste dans V. Dans les deux cas, on procède ensuite au mélange des urnes de la manière précisée page 12.

En revenant au problème général de Markoff, on peut alors se demander ce qui arrive dans le cas régulier pour les $\varpi_j^{(n)}$. On voit qu'alors $\varpi_j^{(n)}$ tendra vers une limite égale à

$$\sum_h P_j \varpi_h = P_j.$$

Dès lors, on en conclut que *dans le cas régulier*, la loi de probabilité définie par les $\varpi_j^{(n)}$ tend vers une loi-limite définie par les P_j, qui est par suite *indépendante de la loi de probabilité initiale définie par les* ϖ_j.

Les limites P_j vérifient, comme on l'a vu, page 47, les équations

$$P_j = \sum_n P_h P_{hj}^{(n)}$$

On a prouvé, que, dans le cas régulier, les équations

$$(\mathcal{E}) \qquad x_j = \sum_h x_h p_{hj}, \qquad (h = 1, \ldots, r); \qquad \sum_j x_j = 1$$

n'ont qu'un système de solutions, qui sont les P_j. Ce résultat exprime qu'il y a alors une loi des probabilités initiales ϖ_j, telle que $\varpi_j^{(1)} = \varpi_j$ et, par suite, $\varpi_j^{(n)} = \varpi_j$ quels que soient les entiers n et j ($\leq r$). On dira qu'une telle loi de probabilité est *stable*. Donc :

Dans le cas régulier, il y a parmi les lois de probabilités initiales (définies par l'expression de ϖ_j en fonction de j) *une loi stable et une seule* et c'est précisément *la loi de probabilité finale* définie par les P_j limites des $P_{hj}^{(n)}$.

Par exemple, dans le cas du mélange des urnes, la loi de probabilité $\varpi_j^{(n)}$ de la composition des urnes après n tirages des urnes U et V varierait avec n si l'on garnissait d'abord U et V comme il a été expliqué avec la troisième urne. D'après le résultat de la page 51, elle

serait au contraire, stable, c'est-à-dire indépendante de n, si l'on opérait d'abord avec la quatrième.

Mais la loi de probabilité initiale peut être stable sans qu'on soit dans le cas régulier. C'est par exemple ce qui a lieu pour $r = 2$, si

$$p_{11} = p_{22} = 0, \qquad p_{12} = p_{21} = 1, \tag{13}$$

d'où

$$P_{ik}^{(n)} = \begin{cases} p_{ik} & \text{si } n \text{ est impair,} \\ 1 - p_{ik} & \text{si } n \text{ est pair,} \end{cases} \tag{14}$$

et si l'on choisit en outre

$$\varpi_1 = \varpi_2 = \frac{1}{2},$$

d'où

$$\varpi_j^{(n)} = \sum_h \varpi_h P_{hj}^{(n)} = \frac{1}{2} \sum_h P_{hj}^{(n)} = \frac{1}{2} = \varpi_j$$

Par contre, il ne suffit pas qu'on soit dans le cas régulier pour qu'une loi de probabilité initiale soit nécessairement stable. Par exemple, si l'on prend $\varpi_h = 0$ sauf pour $h = 1$, on a $\varpi_j^{(n)} = P_{1j}^{(n)}$ et l'on sait bien qu'en général $P_{1j}^{(n)}$ n'est pas indépendant de n.

En revenant au cas général, notons pour la suite qu'on a toujours, en vertu de (12),

$$\sum_j \varpi_j^{(n)} = 1, \tag{15}$$

et, d'autre part, d'après (12) et (I),

$$\varpi_j^{(n+t)} = \sum_i \varpi_i^{(n)} P_{ij}^{(t)}. \tag{16}$$

Probabilités en chaînes inverses. — M. S. Bernstein [5] en 1932, M. Mihoc [1] en 1934, puis, M. Hostinský [15] et M. Kolmogoroff [2, 3] à peu près simultanément, ont commencé à étudier les probabilités inverses des événements en chaîne. Un peu après, M. Potoček [2] et M. Onicescu [2] s'en sont également occupés. Nous allons exposer leurs résultats en les complétant sur certains points. Il s'agit d'abord d'étudier la probabilité $q_{jk}^{(n)}$ pour que le système considéré se soit trouvé à la $(n-1)^{\text{ième}}$ épreuve à l'état E_j

quand, à la $n^{\text{ième}}$ épreuve, il s'est trouvé a l'état E_k. Calculons $q_{jk}^{(n)}$. D'après le théorème des probabilités composées, la probabilité pour que se produisent successivement les deux événements suivants : le système considéré est à l'état E_j à la $(n-1)^{\text{ième}}$ épreuve et à l'état E_k à la $n^{\text{ième}}$, peut se calculer de deux façons qui entraînent l'égalité

$$\varpi_j^{(n-1)} p_{jk} = \varpi_k^{(n)} q_{jk}^{(n)} \tag{17}$$

Si donc $\varpi_k^{(n)} \neq 0$, on a

$$q_{jk}^{(n)} = p_{jk} \frac{\varpi_j^{(n-1)}}{\varpi_k^{(n)}} \tag{18}$$

Si $\varpi_k^{(n)} = 0$, $q_{jk}^{(n)}$ désignerait la probabilité d'un événement se produisant quand se produit un autre événement (E_k à la $n^{\text{ième}}$ épreuve) qui est impossible ou au moins dont la probabilité est nulle . $q_{jk}^{(n)}$ n'a plus alors une signification précise et il n'y a pas lieu de le calculer

Notons, d'après ce qui précède, que

$$q_{jk}^{(n)} = p_{jk} \frac{\sum_i \varpi_i P_{ij}^{(n-1)}}{\sum_i \varpi_i P_{ik}^{(n)}},$$

et que d'après la signification de $q_{jk}^{(n)}$, ou d'après cette égalité, on a

$$\sum_j q_{jk}^{(n)} = 1. \tag{19}$$

On observe que si la loi de probabilité initiale ϖ_j est stable, la formule (17) se réduit à

$$\varpi_k q_{jk}^{(n)} = \varpi_j p_{jk}, \tag{20}$$

de sorte que dans ce cas, si la loi initiale est « positivement » stable, c'est-à-dire, si, en outre, les ϖ_k sont tous $\neq 0$, $q_{jk}^{(n)}$ devient indépendant de n, on peut alors le désigner par q_{jk} et l'on a

$$q_{jk} = p_{jk} \frac{\varpi_j}{\varpi_k}. \tag{21}$$

Réciproquement, si $q_{jk}^{(n)}$ est indépendant de n (soit $q_{jk}^{(n)} = q_{jk}$), la loi de probabilité initiale est stable sauf en un cas d'exception, celui où l'un

des couples p_{jk}, q_{jk} est formé de zéros. En effet, considérons une valeur de k pour laquelle tous les p_{jk} sont $\neq 0$. on a (15) et

$$\varpi_j^{(n-1)} = \frac{q_{jk}}{p_{jk}} \varpi_k^{(n)}, \qquad \text{d'où} \quad \varpi_k^{(n)} \sum_j \frac{q_{jk}}{p_{jk}} = 1$$

et, par suite, $\varpi_k^{(n)}$ est indépendant de n et différent de zéro.

Si, au contraire, pour k donné, l'un des p_{jk} est nul, on aurait $\varpi_k^{(n)} q_{jk} = 0$ et $\varpi_k q_{jk} = 0$ Alors ou bien $\varpi_k^{(n)} = 0$ quel que soit n et par suite $\varpi_k^{(n)}$ est encore indépendant de n, ou bien $q_{jk} = 0$

Dans le cas où, pour au moins un système d'indices j, k, on a $p_{jk} = 0$ et $q_{jk} = 0$ la loi de probabilité initiale peut ne pas être stable.

Plaçons-nous par exemple dans le cas des formules (13), mais avec $\varpi_1 \neq \varpi_2 = 1 - \varpi_1$, $\varpi_1 \neq 0$, $\varpi_2 \neq 0$ Alors, en vertu de (12),

$$\varpi_j^{(n)} = \varpi_1 P_{1j}^{(n)} + \varpi_2 P_{2j}^{(n)}$$

est egal alternativement quand n croît à ϖ_2 et ϖ_1, avec $\varpi_2^{(n)} = 1 - \varpi_1^{(n)}$ On a aussi

$$0 = \varpi_j^{(n-1)} p_{jj} = q_{jj}^{(n)} \varpi_j^{(n)}$$

avec $\varpi_j^{(n)} \neq 0$, d'où $q_{jj}^{(n)} = 0$ et d'après (19) $q_{jk}^{(n)} = 1$ pour $j \neq k$ Donc les $q_{jk}^{(n)}$ sont independants de n Pourtant, on vient de voir que, $\varpi_j^{(n)}$ prenant alternativement des valeurs distinctes, la loi des ϖ_j n'est pas stable.

Observons que les cas d'exception disparaissent, lorsque tous les p_{ik} sont $\neq 0$. *dans ce cas la condition nécessaire et suffisante pour que les probabilités inverses $q_{jk}^{(n)}$ soient indépendantes de n est que la loi de probabilité initiale soit positivement stable.*

Dans ce cas la formule (21) peut s'écrire :

$$q_{jk} = p_{jk} \frac{P_j}{P_k}, \tag{22}$$

d'où, en particulier,

$$q_{jj} = p_{jj}.$$

En vertu de (19) et de (T_1), il en résulte que, pour $r = 2$, on a, dans le cas stable et positivement régulier.

$$q_{kj} = p_{jk} \tag{23}$$

que j et k soient égaux ou non.

Enfin dans le cas stable « le plus régulier », on a d'apres (22), quels que soient r, j, k,

$$q_{jk} = p_{jk} \tag{24}$$

Considérons maintenant plus généralement, la probabilité $Q_{jk}^{(m,n)}$ pour que l'état E_k étant réalisé à la $n^{\text{ième}}$ épreuve, le système ait été à l'état E_j à l'épreuve antérieure de rang $m\,(m < n)$. On aura, en écrivant de deux façons la valeur de la probabilité que les états E_j, E_k se soient présentés à la $m^{\text{ième}}$ et à la $n^{\text{ième}}$ épreuve respectivement :

$$\varpi_j^{(m)} P_{jk}^{(n-m)} = \varpi_k^{(n)} Q_{jk}^{(m,n)} \tag{25}$$

Si s est entre m et n $(m < s < n)$, on aura

$$\varpi_k^{(n)} Q_{jk}^{(m,n)} = \sum_i \varpi_j^{(m)} P_{ji}^{(s-m)} P_{ik}^{(n-s)} = \sum_i \varpi_i^{(s)} Q_{ji}^{(m,s)} P_{ik}^{(n-s)} = \sum_i Q_{ji}^{(m,s)} \varpi_k^{(n)} Q_{ik}^{(s,n)};$$

par suite, si $\varpi_k^{(n)} \neq 0$, on a l'égalité d'itération

$$Q_{jk}^{(m,n)} = \sum_i Q_{ji}^{(m,s)} Q_{ik}^{(s,n)} \qquad \text{pour} \quad m < s < n \tag{26}$$

Comme $Q_{jk}^{(n-1,n)} = q_{jk}^{(n)}$, on voit d'après cette égalité que connaissant les $q_{ij}^{(m+1)}, q_{il}^{(m+2)}, \ldots, q_{jk}^{(n)}$ on peut calculer de proche en proche $Q_{jk}^{(m,n)}$.

Observons aussi que, *dans le cas positivement stable*, on a

$$Q_{jk}^{(m,n)} = \frac{\varpi_j}{\varpi_k} P_{jk}^{(n-m)},$$

de sorte qu'alors $Q_{jk}^{(m,n)}$ *ne dépend de m et n que par l'intermédiaire de* $n - m$, et l'on peut poser

$$Q_{jk}^{(n-m)} = Q_{jk}^{(m,n)}.$$

Dans ce cas, en posant $s - m = \alpha$, $n - s = \beta$, l'égalité (26) prend la forme plus simple

$$Q_{jk}^{(\alpha+\beta)} = \sum_i Q_{ji}^{(\alpha)} Q_{ik}^{(\beta)}, \tag{27}$$

équation d'itération entièrement analogue à celle qui régit les $P_{ik}^{(\alpha)}$, mais qui n'a été établie qu'en supposant stable la loi de probabilité initiale des ϖ_j. On aura dans ce cas $Q_{jk}^{(1)} = q_{jk}$. Si, en outre, on se

trouve dans le cas positivement régulier, on aura

$$Q_{jk}^{(n)} = \frac{P_j}{P_k} P_{jk}^{(n)}. \tag{28}$$

Qu'on soit ou non dans le cas stable, on voit, d'après (25), que dans le cas positivement régulier, $Q_{jk}^{(m,n)}$ tend vers une limite $Q_{jk}^{(m,\infty)}$ lorsque m restant fixe, n croît indéfiniment. Et cette limite, égale à $\varpi_j^{(m)}$, est indépendante de l'état final E_k. Si n, et non seulement m, mais aussi $n-m$ tendent simultanément vers l'infini, $Q_{jk}^{(m,n)}$ tend vers P_j.

Nous venons d'exposer avec quelques compléments les résultats de MM. Hostinský et Potoček. Avant de passer à ceux de M. Kolmogoroff, faisons encore une observation.

On peut généraliser l'égalité (24) et voir que, dans le cas régulier, la condition nécessaire et suffisante pour que l'on ait, quels que soient j, k, m, n $(m<n)$,

$$Q_{jk}^{(m,n)} = P_{jk}^{(n-m)} \qquad (\text{donc en particulier } q_{jk} = p_{jk}), \tag{28}$$

est qu'on soit dans le cas « le plus régulier » [c'est-à-dire que la condition (T'_1) soit vérifiée] et qu'en outre la loi de probabilité initiale soit stable. D'après (24) cette double condition suffit pour qu'on ait $q_{jk}=p_{jk}$, et d'après les formules d'itération (I) et (26), cette dernière égalité suffit aussi pour assurer (28).

Pour démontrer que cette double condition est nécessaire, observons que si (28) est vérifié, on a, d'après (25),

$$[\varpi_k^{(n)} - \varpi_j^{(m)}] P_{jk}^{(n-m)} = 0. \tag{29}$$

Faisons croître m et $n-m$ indéfiniment; on aura

$$(P_k - P_j) P_k = 0.$$

En vertu de $\sum_i P_i = 1$, l'un au moins des P_i, par exemple P_k est $\neq 0$. On aura donc quel que soit j, $P_j = P_k$; on est bien dans le cas le plus régulier.

Laissons maintenant k et m fixes. Pour n assez grand $(n-m>A_j)$, on a $P_{jk}^{(n-m)} \neq 0$; donc, d'après (29),

$$\varpi_k^{(n)} = \varpi_j^{(m)}$$

pour $n - m > A_j$, et en appelant A le plus grand des A_j

$$\varpi_j^{(m)} = \varpi_k^{(n)} \tag{30}$$

quel que soit j pour m fixe et $n - m > A$. Ainsi, quel que soit m,

$$\varpi_1^{(m)} = \varpi_2^{(m)} = \ldots = \varpi_r^{(m)}$$

Par suite, on a, quels que soient m et j,

$$\varpi_j^{(m)} = \frac{1}{r}.$$

Donc la loi de probabilité initiale est stable.

Cette loi stable est d'ailleurs très simple. Puisqu'on est dans le cas stable le plus régulier, on doit avoir ici $\varpi_j = P_j = \frac{1}{r}$, c'est-à-dire les ϖ_j sont égaux, c'est-à-dire que la loi de probabilité initiale doit être « uniforme ».

Suite d'épreuves illimitée dans les deux sens. — M. Kolmogoroff [2], puis MM. Onicescu [2] et Mihoc ont aussi considéré le cas fréquent dans la nature, où l'on ne peut assigner une origine à la suite d'épreuves, c'est-à-dire où l'on peut pratiquement la supposer illimitée dans les deux sens.

Probabilités absolues. — Alors il n'y a plus une loi de probabilité proprement initiale, mais il peut y avoir encore une probabilité $\varpi_j^{(n)}$ pour que le système se trouve à l'état E_j à la $n^{\text{ième}}$ épreuve, n pouvant, cette fois, être positif, négatif ou nul. On pourra appeler avec M. Kolmogoroff les $\varpi_j^{(n)}$, les probabilités absolues et les p_{jk}, $P_{jk}^{(n)}$ les probabilités de passage, et l'on aura encore les équations (15), (16), (17), (25) et (26).

Quand les $\varpi_j^{(n)}$ existent, ils sont liés par la relation de récurrence (16). C'est donc le cas où les probabilités p_{jk} étant données — donc aussi les $P_{jk}^{(t)}$ pour $t > 0$ — il y a au moins un système de solutions $x_i^{(n)} = \varpi_i^{(n)}$ du système de relations

$$(\mathcal{S}') \qquad x_j^{(n+t)} = \sum_{i=1}^{i=r} x_i^{(n)} P_{ij}^{(t)}, \qquad \sum_{i=1}^{i=r} x_i^{(n)} = 1 \qquad (x_i^{(n)} \geqq 0),$$

où n, t sont entiers, $t \geqq 0$, n de signe quelconque.

M. Kolmogoroff [2, p. 156] a même prouvé que, quel que soit le choix des probabilités p_{jk} $\left(\text{telles que } p_{jk} \geqq 0, \ \sum_{k=1}^{k=r} p_{jk} = 1\right)$, le système ($\mathcal{E}'$) a un système de solutions au moins. Il a ensuite cherché la condition d'unicité des solutions.

Sa démonstration est valable dans un cas plus général que le cas actuel. On va ici, en se limitant à ce cas particulier, donner une autre démonstration pour la condition nécessaire en utilisant certains résultats non limités au cas régulier et démontrés plus loin.

On pourrait d'abord observer qu'il suffit de remplacer dans le système $x_j^{(n)}$ par $P_{hj}^{(n)}$ pour constater que ce système est vérifié. Mais $P_{hj}^{(n)}$ n'étant défini que pour n (entier) positif ne fournit pas une solution complète.

Au contraire, si l'on admet que, comme il sera démontré page 259, $P_{hj}^{(n)}$ converge toujours « en moyenne arithmétique » (voir p. 71) vers une quantité Π_{hj}, on observe que le système ($\mathcal{E}'$) est vérifié pour $x_j^{(n)} = \Pi_{hj}$, comme on le verra page 261, et ceci pour n d'un signe quelconque.

Nous avons ainsi r systèmes de solutions pour $h = 1, \ldots, r$.

Mais elles sont indépendantes de n. On peut en trouver qui, en général, dépendent de n, en introduisant la période asymptotique, N, de $P_{ij}^{(1)}, P_{hj}^{(2)}, \ldots$ définie plus loin page 260.

Quand s croît, $P_{hi}^{(n+sN)}$ tend, comme on le verra page 113, vers une limite $W_{hi}^{(n)}$ qui ne change pas quand on remplace n par un autre nombre n' tel que $n - n'$ soit un multiple de N. Tout ceci suppose n, n' positifs. Mais il est clair que $W_{hi}^{(n)}$ étant défini pour n entier > 0, on peut poser par définition $W_{hi}^{(n)} = W_{hi}^{(n+\sigma N)} = W_{hi}^{(\alpha)}$ pour n entier quelconque, où σ est un entier de signe quelconque tel que $\alpha = n + \sigma N$ soit l'un des nombres $1, \ldots, N$.

On a évidemment

$$W_{hi}^{(n)} \geqq 0, \qquad \sum_i W_{hi}^{(n)} = 1.$$

Faisons croître s dans la relation

$$P_{hj}^{(n+sN+t)} = \sum_i P_{hi}^{(n+sN)} P_{ij}^{(t)}.$$

On aura

$$W_{hj}^{(n+t)} = \sum_{i} W_{hi}^{(n)} P_{ij}^{(t)}.$$

Ainsi on a encore r systèmes de solutions du système ($\mathcal{E}'$), soient les solutions $x_j^{(n)} = W_{hj}^{(\alpha)}$.

Ces solutions ne sont d'ailleurs pas nécessairement distinctes.

Le cas le plus intéressant est celui où le système ($\mathcal{E}'$) n'a qu'un seul système de solutions. Il faut au moins pour cela que les r solutions Π_{hj} se confondent : $\Pi_{hj} = \Pi_j$, c'est-à-dire qu'on soit au moins dans le cas semi-régulier.

Mais il faut aussi que les solutions $W_{hj}^{(n)}$ et Π_j coïncident, que par conséquent $W_{hi}^{(n)}$ soit indépendant de n, et par suite que $W_{hi}^{(1)} = \ldots = W_{hi}^{(N)}$, ce qui exige qu'on soit au moins dans le cas non oscillant. Il faut donc, en définitive, qu'on soit dans le cas régulier

Réciproquement, cette condition est suffisante, comme le montre le calcul de M. Kolmogoroff. On a, en effet, pour toute solution $X_j^{(n)}$ du système ($\mathcal{E}'$) (nous savons qu'il en existe au moins une),

$$|X_j^{(n)} - P_j| = \left| \sum_{i} X_i^{(-t)} [P_{ij}^{(n+t)} - P_j] \right| \leqq \sum_{i=1}^{i=r} |P_{ij}^{(n+t)} - P_j|,$$

où, quel que soit le signe de n, on peut prendre t assez grand pour que $n + t$ soit > 0 et par suite que $P_{ij}^{(n+t)}$ ait un sens. Faisons croître t, le dernier terme tend par hypothèse vers zéro. Donc $X_j^{(n)} = P_j$. Il n'y a qu'une solution. Nous voyons même que cette solution est indépendante de n.

Si, maintenant, nous supposons qu'il y a un système de probabilités absolues $\varpi_j^{(n)}$, nous voyons que, dans le cas régulier, ce système est nécessairement stable et qu'on a nécessairement $\varpi_j^{(n)} \equiv P_j$, où $P_j = \lim\limits_{n \to \infty} P_{ij}^{(n)}$, résultat qui n'était pas valable dans le cas examiné plus haut d'une probabilité initiale.

Probabilités inverses. — On a encore la formule (17) quand la suite des épreuves est illimitées dans les deux sens. Mais dans le cas régulier, les probabilités absolues sont indépendantes de n, et (17) peut s'écrire

$$P_j p_{jk} = P_k q_{jk}^{(n)}.$$

Des lors, on voit que dans le cas positivement régulier les probabilités inverses sont stables : $q_{jk}^{(n)} = q_{jk}$, et l'on a

$$q_{jk} = \frac{P_j p_{jk}}{P_k}.$$

En particulier, dans le cas le plus régulier et seulement dans ce cas, on a

$$q_{jk} = p_{jk}$$

On a aussi

$$(25) \qquad \varpi_j^{(m)} P_{jk}^{(n-m)} = \varpi_k^{(n)} Q_{jk}^{(m,n)},$$

de sorte que dans le cas positivement régulier les $Q_{jk}^{(m,n)}$ ne dépendent de m et n que par l'intermédiaire de $n-m$, et l'on a

$$(32) \qquad Q_{jk}^{(m,n)} = Q_{jk}^{(n-m)} = \frac{P_j P_{jk}^{(n-m)}}{P_k},$$

d'où, en particulier,

$$Q_{jk}^{(\alpha+\beta)} = \sum_i Q_{ji}^{(\alpha)} Q_{ik}^{(\beta)}$$

Dans le cas le plus régulier, on a

$$(33) \qquad Q_{jk}^{(\alpha)} = P_{jk}^{(\alpha)}$$

Et réciproquement, cette égalité ne peut avoir lieu dans le cas positivement régulier que si l'on se trouve en même temps dans le cas le plus régulier.

On peut aussi se demander avec M. Kolmogoroff [2, p. 158] dans quel cas on aura

$$(34) \qquad Q_{kj}^{(\alpha)} = P_{jk}^{(\alpha)}.$$

Ceci exprime que la probabilité d'avoir été à une épreuve, de rang arbitraire déterminé n, à l'état E_k quand on sait qu'on est à l'état E_j à l'épreuve de rang $n+\alpha$ est égale à la probabilité d'être à l'état E_k à l'épreuve de rang $n+\alpha$ quand on sait qu'on a été à l'état E_j à l'épreuve de rang n.

Lorsqu'on se trouve dans le cas positivement régulier, la condition est évidemment, d'après (32), que l'on ait

$$(35) \qquad P_k P_{kj}^{(\alpha)} = P_j P_{jk}^{(\alpha)},$$

c'est-à-dire que la quantité $\theta_{jk}^{(\alpha)} = \mathrm{P}_j \mathrm{P}_{jk}^{(\alpha)}$ soit une fonction symétrique de j et de k. Un cas particulier important où il en est ainsi est le cas symétrique où les p_{jk} sont symétriques en j et k. Car alors, il en est de même de $\mathrm{P}_{jk}^{(\alpha)}$ et, d'autre part, on a vu (p. 40) que dans ce cas les P_j sont égaux.

Pour le cas général, on a, en prenant $\alpha = 1$,

$$\mathrm{P}_k p_{kj} = \mathrm{P}_j p_{jk}$$

On peut alors en déduire avec M. Kolmogoroff [2] une forme de la condition (35) qui ne nécessite pas le calcul des P_k.

En effet, en écrivant que

$$\theta_{ab}\theta_{bc}\ldots\theta_{ij}\theta_{jk}\theta_{ka} = \theta_{ak}\theta_{kj}\theta_{ji}\ldots\theta_{cb}\theta_{ba},$$

les P_h qui sont supposés $\neq 0$ s'éliminent et, pour $\alpha = 1$, il reste la condition

$$p_{ab}p_{bc}\ldots p_{ij}p_{jk}p_{ka} = p_{ak}p_{kj}p_{ji}\ldots p_{cb}p_{ba} \tag{36}$$

qui doit être vérifiée en prenant pour les $a, b, \ldots, k$ des entiers quelconques de 1 à r. Réciproquement, si cette condition est vérifiée, on a

$$\sum_{b,c,\ldots,f} p_{ab}p_{bc}\ldots p_{fg} \sum_{h,i,\ldots,k} p_{gh}p_{hi}\ldots p_{ka} = \sum_{k,i,\ldots,h} p_{ak}p_{ki}\ldots p_{hg} \sum_{f,\ldots,b} p_{gf}\ldots p_{ba}$$

ou

$$\mathrm{P}_{ag}^{(\beta)}\mathrm{P}_{ga}^{(\alpha)} = \mathrm{P}_{ag}^{(\alpha)}\mathrm{P}_{ga}^{(\beta)}.$$

En faisant croître β, on a donc

$$\mathrm{P}_g \mathrm{P}_{ga}^{(\alpha)} = \mathrm{P}_a \mathrm{P}_{ag}^{(\alpha)},$$

c'est-à-dire la condition (35). Ainsi, dans le cas positivement régulier, la condition nécessaire et suffisante pour qu'on ait la formule de réversibilité (34) est la condition (36). Cette formule exprime que la probabilité de passer successivement par des états déterminés et de revenir à l'état initial est la même quand leur ordre forme une permutation circulaire ou la permutation circulaire de sens contraire.

b — **Étude d'une variable aléatoire « en chaîne » dans le cas régulier.**

Valeur moyenne. — Supposons qu'une variable aléatoire $X(E)$ soit déterminée par le résultat fortuit, E, d'une épreuve, et que celle-ci amène un certain Système S à l'un, E_k, par exemple, des états possibles de S. A E_k correspond une valeur déterminée x_k de $X(E)$. Ainsi E_l peut désigner la situation d'une carte occupant le rang l dans un paquet, et l'on peut prendre pour $x_l = X(E_l)$ une fonction déterminée de ce rang, son carré par exemple. Si les probabilités des résultats d'épreuves successives sont « en chaîne », alors nous pourrons désigner par $Y_h^{(n)} = Y^{(n)}(E_h) = X(E_l) = x_l$, la valeur prise par $X(E)$ quand E_l est le résultat de n épreuves à partir de E_h. Quand n et h sont donnés, la valeur x_l de $Y_h^{(n)}$ n'est fixée que par le hasard, mais sa probabilité étant $P_{hl}^{(n)}$, on peut en calculer la valeur moyenne ([1])

$$\mathfrak{M} Y_h^{(n)} = \overline{Y}_h^{(n)} = \sum_l P_{hl}^{(n)} x_l, \tag{37}$$

qui, elle, sera déterminée, connaissant h et n.

On observe qu'en vertu des relations d'itération, on a

$$\mathfrak{M} Y_h^{(n+m)} = \sum_j P_{hj}^{(n)} \mathfrak{M} Y_j^{(m)}.$$

Autrement dit, $Y_h^{(n)}$ s'obtient par l'itération répétée n fois de la transformation

$$v_k = \sum_{j=1}^{j=r} p_{kj} u_j \qquad (k = 1, \ldots, r)$$

à partir des valeurs $u_j = x_j$.

Quand nous aurons examiné, plus tard, le comportement de $P_{hl}^{(n)}$ dans le cas le plus général, nous pourrons examiner (p. 132) celui de $\mathfrak{M} Y_h^{(m)}$ avec la même généralité. Pour le moment, nous allons nous borner à examiner ce qui se passe *dans le cas régulier*.

([1]) Suivant la commodité, nous représenterons la valeur moyenne — au sens du Calcul des Probabilités — d'une variable aléatoire Z, tantôt par $\mathfrak{M} Z$, tantôt par $\overline{Z}$.

On a alors

$$\lim_{n \to \infty} P_{hk}^{(n)} = P_k,$$

et, par suite,

$$\lim_{n \to \infty} \mathfrak{M} Y_h^{(n)} = \sum_l P_l x_l = M$$

Cette limite M est indépendante de h : *dans le cas régulier, la valeur moyenne du nombre aléatoire* X(E) *après n épreuves tend, lorsque n croît indéfiniment, vers une limite indépendante de sa valeur initiale.*

On a vu (p. 45) que dans le cas régulier, le système des équations qui déterminent les P_k

$$P_k = \sum_j P_j p_{jk}, \qquad \sum_j P_j = 1$$

n'a qu'un seul système de solutions. Alors, on pourra obtenir la limite M en fonction des données en éliminant $P_1, \ldots, P_r$ entre ces équations et l'équation

$$M = \sum_j P_j x_j \tag{38}$$

Il est d'ailleurs bien clair que *cette limite* M de la valeur moyenne de $Y_h^{(n)}$ *est elle-même une valeur moyenne*, à savoir celle d'une quantité aléatoire Z dont les seules valeurs possibles sont celles, $x_1, \ldots, x_r$, de X, mais avec des probabilités $P_1, \ldots, P_r$ qui sont les limites de $P_{h1}^{(n)}, \ldots, P_{hr}^{(n)}$. Pour évidente qu'elle paraisse, cette propriété de la limite des valeurs moyennes d'être une valeur moyenne cessera d'avoir toujours lieu dans le cas où le nombre des états possibles est infini.

Remarque I. — Il n'y a, lorsqu'on porte son attention sur les valeurs moyennes, rien qui distingue le cas positivement régulier du cas régulier général. Il importe peu dans cette question que l'une ou l'autre des limites P_k soit nulle, puisque dans l'expression de M, les P_k sont multipliés par des quantités x_k qui sont chacune > 0, < 0 ou $= 0$. $\left(\text{Les } P_k \text{ ne sont d'ailleurs jamais toutes nulles, car } \sum P_k = 1 \text{ et } P_k \geqq 0.\right)$

Remarque II. — Dans le cas le plus régulier, celui où la condition (T'_1) est vérifiée, les limites P_k ayant une même valeur $P = \frac{1}{r}$, la limite M de la valeur moyenne $Y_h^{(n)}$ se réduit *à la moyenne arithmétique*

$$M = \frac{x_1 + x_2 + \dots + x_r}{r}$$

des valeurs possibles de X(E).

III. La quantité $[\mathfrak{M} Y_h^{(n)} - M]$ est un infiniment petit avec $\frac{1}{n}$; on peut même préciser. Car, dans l'expression

$$\mathfrak{M} Y_h^{(n)} - M = \sum_{l=1}^{l=r} x_l [P_{hl}^{(n)} - P_l],$$

les crochets sont majorés par les termes d'une progression géométrique convergente (p. 31). $[\mathfrak{M} Y_h^{(n)} - M]$ est donc un infiniment petit *d'un ordre supérieur* à σ^n où σ est un certain nombre > 0 et < 1.

Application au mélange des urnes (p. 12, 34). — Dans ce problème, prenons pour $Y_h^{(n)}$, la proportion du nombre des boules blanches dans l'urne U au bout de n opérations quand il y en avait primitivement h. Alors $u Y_h^{(n)}$ sera le nombre de ces boules blanches. Ce nombre peut prendre les valeurs 0, 1, ..., r (r étant le plus petit des nombres u et B) avec des probabilités $P_{h0}^{(n)}, \dots, P_{hr}^{(n)}$ qui dépendent du nombre initial h des boules blanches dans U. On a donc

$$u \mathfrak{M} Y_h^{(n)} = \sum_{k=0}^{k=r} k P_{hk}^{(n)}$$

On voit d'abord que $\mathfrak{M} Y_h^{(n)}$ tend vers une limite

$$M = \sum_{k=0}^{k=r} \frac{k}{u} P_k = \sum_{k=1}^{k=r} \frac{k}{u} P_k$$

qui est indépendante de la composition initiale des urnes U et V. De

plus, d'après les formules (10) de la page 51, on a ici

$$\frac{k}{u} P_k = \frac{k}{u} \frac{C_u^k C_v^{B-k}}{C_T^B} = \frac{C_{u-1}^{k-1} C_v^{B-k}}{C_T^B},$$

d'où

$$M = \sum_{k=1}^{k=r} \frac{C_{u-1}^{k-1} C_v^{B-k}}{C_T^B}.$$

D'ailleurs, l'identité $\sum_{k=0}^{k=r} P_k = 1$ se traduit par l'identité

$$\sum_{k=0}^{k=r} C_u^k C_v^{B-k} = C_{u+v}^B \quad (^1).$$

En y remplaçant u par $u-1$, B par $B-1$, donc r par $r-1$ et k par $k-1$, on a

$$\sum_{k=1}^{k=r} C_{u-1}^{k-1} C_v^{B-k} = C_{u-1+v}^{B-1},$$

identité qui, portée, dans l'expression de M fournit

$$\lim_{n \to \infty} \mathfrak{M} Y_k^{(n)} = \frac{C_{T-1}^{B-1}}{C_T^B} = \frac{B}{T}.$$

C'est le théorème démontré par Laplace la valeur moyenne de la proportion des boules blanches dans l'urne U après n opérations tend, lorsque n croît indéfiniment, vers la proportion des boules blanches dans une urne où l'on aurait versé U et V.

Le même résultat ayant lieu pour l'urne V, on peut dire simplement que *les proportions moyennes des boules blanches dans les deux urnes tendent à s'égaliser*.

Seulement, ici le résultat est établi *dans un cas plus général :* il n'a pas été nécessaire de supposer avec Laplace que les nombres de boules de chaque couleur et de chaque urne étaient très grands.

Moyennes arithmétiques. — En revenant au problème général de

(1) Notons en passant cet exemple de formules d'Analyse combinatoire démontrées par le détour du Calcul des Probabilités.

Markoff, on peut aussi donner une autre interprétation de M en faisant intervenir la moyenne *arithmétique* $A_h^{(n)}$ de $Y_h^{(1)}, \ldots, Y_h^{(n)}$ et la valeur moyenne $\mathfrak{M}A_h^{(n)}$ de $A_h^{(n)}$, soit

$$(38\ bis)\qquad M_h^{(n)} = \mathfrak{M}A_h^{(n)} = \frac{1}{n}[\mathfrak{M}Y_h^{(1)} + \mathfrak{M}Y_h^{(2)} + \ldots + \mathfrak{M}Y_h^{(n)}]$$

On sait (¹) que si une suite de nombres $u_1, u_2, \ldots$ tend vers une limite λ, il en est de même de la suite des moyennes arithmétiques des premiers nombres u

$$\lim_{n\to\infty} \frac{u_1 + u_2 + \ldots + u_n}{n} = \lambda.$$

D'ailleurs, on sait que la réciproque de la proposition rappelée ci-dessus n'est pas exacte. Par exemple, en prenant $u_n = (-1)^n$, la moyenne arithmétique des n premiers nombres u_k tend vers zéro, tandis que les u_k n'ont pas de limite. On peut donc dire que la limite de cette moyenne quand elle existe est une limite généralisée, qui se confond avec la limite usuelle quand celle-ci existe.

Or, $\mathfrak{M}Y_h^{(n)}$ tend vers M; on voit donc qu'on a

$$(39)\qquad \lim_{n\to\infty} \mathfrak{M}A_h^{(n)} = M.$$

L'intérêt de la formule (39) réside précisément dans le fait qui sera démontré plus loin (p. 132) que $\mathfrak{M}A_h^{(n)}$ garde encore une limite (mais qui pourra dépendre de h) même dans le cas singulier, même, en particulier, quand $\mathfrak{M}Y_h^{(n)}$ n'a pas de limite déterminée [*voir* note (¹), p. 141].

Remarque. — On vient de prouver que dans le cas régulier $M_h^{(n)} - M$ est infiniment petit avec $\frac{1}{n}$. On peut préciser. On a, en effet,

$$n[M_h^{(n)} - M] = \sum_{t=1}^{t=n} [\overline{Y}_h^{(t)} - M]$$
$$= \sum_{t=1}^{t=n} \sum_j [P_{hj}^{(t)} - P_j]x_j = \sum_j \left\{ \sum_{t=1}^{t=n} [P_{hj}^{(t)} - P_j] \right\} x_j.$$

(¹) Voir, par exemple, M. Fréchet, *Leçons sur les séries trigonométriques* (*Centre de documentation universitaire*, 1935, p. 23).

En appelant s_{hj}, comme à la page 31, la somme de la série absolument convergente

$$s_{hj} = \sum_{l=1}^{l=\infty} [P_{hj}^{(l)} - P_j],$$

on a donc

$$(40) \qquad \lim_{n \to \infty} n(M_h^{(n)} - M) = S_h = \sum_{j=1}^{j=\nu} s_{hj}\, x_j.$$

Ainsi, $M_h^{(n)} - M$ est un infiniment petit d'ordre au moins égal a celui de $\frac{1}{n}$; si $S_h \neq 0$, sa partie principale est $\frac{S_h}{n}$, si $S_h = 0$, son ordre est supérieur à celui de $\frac{1}{n}$.

On a d'ailleurs vu plus haut (p. 47) qu'il n'est pas nécessaire pour calculer les s_{hj} et par suite S_h de calculer toutes les probabilités itérées $P_{hj}^{(n)}$.

Remarque. — D'apres (12) on a

$$M_h^{(n)} = \sum_j x_j \Pi_{hj}^{(n)},$$

en posant

$$(41) \qquad \Pi_{hj}^{(n)} = \frac{1}{n}(P_{hj}^{(1)} + P_{hj}^{(2)} + \ldots + P_{hj}^{(n)}).$$

Nous allons donner une interprétation de $\Pi_{hj}^{(n)}$.

Fréquence moyenne. — Il est intéressant de calculer ici, comme dans le cas des probabilités constantes de Bernoulli, la moyenne de la fréquence d'un événement. Précisons.

Il y a plusieurs fréquences à distinguer.

D'abord, la fréquence $F_{hk}^{(n,N)}$ avec laquelle, dans N groupes de n épreuves, on trouve l'état E_k à la $n^{\text{ième}}$ épreuve, après être parti de l'état initial E_h. C'est la fréquence d'un événement de probabilité $P_{hk}^{(n)}$ constante d'un groupe à l'autre. On a évidemment

$$(42) \qquad \mathfrak{M}\, F_{hk}^{(n,N)} = P_{hk}^{(n)}.$$

On voit que, dans le cas régulier, on aura

$$(42\ bis) \qquad \lim_{n \to \infty} \mathfrak{M}\, F_{hk}^{(n,N)} = P_k.$$

Mais on peut aussi considérer une autre fréquence en envisageant ce qui se passe, non à la $n^{\text{ième}}$ épreuve, mais dans l'ensemble des n premières épreuves.

Au cours de n épreuves successives, en partant d'un certain état E_h, un état E_k choisi d'avance peut se présenter, se « répéter » $\rho_{hk}^{(n)}$ fois, $\rho_{hk}^{(n)}$ sera la « répétition » de E_k dans ces n épreuves, et

$$f_{hk}^{(n)} = \frac{\rho_{hk}^{(n)}}{n}$$

en sera la fréquence.

M. von Mises [1, p. 551-553] a calculé la moyenne et l'écart quadratique moyen de $f_{hk}^{(n)}$. Nous suivrons une méthode adoptée fréquemment maintenant pour le cas de Bernoulli et basée sur l'emploi de la notion de variable aléatoire.

Nous considérerons $\rho_{hk}^{(n)}$ comme la somme $\sum_{t=1}^{t=n} U^{(t)}$ de variables aléatoires dont chacune $U^{(t)}$ prend la valeur 1 ou 0 suivant que, partant de l'état E_h, on est arrivé ou non, après t épreuves, à l'état E_k. Donc

$$\mathfrak{M}\,\rho_{hk}^{(n)} = \sum_{t=1}^{t=n} \mathfrak{M}\,U^{(t)}$$

Or, quand $U^{(t)}$ prend la valeur 1, c'est avec la probabilité $P_{hk}^{(t)}$, donc

$$\mathfrak{M}\,U^{(t)} = P_{hk}^{(t)}$$

On a ainsi, sans calcul,

$$\mathfrak{M}\,\rho_{hk}^{(n)} = P_{hk}^{(1)} + \ldots + P_{hk}^{(n)},$$

et, par suite,

$$\mathfrak{M}\,f_{hk}^{(n)} = \Pi_{hk}^{(n)} = \frac{P_{hk}^{(1)} + \ldots + P_{hk}^{(n)}}{n}. \tag{43}$$

Nous trouvons ici l'interprétation annoncée de $\Pi_{hk}^{(n)}$. Comme dans le cas de Poisson : la valeur moyenne de la fréquence de E_k (à partir de E_h) dans n épreuves est égale à la moyenne arithmétique des probabilités d'arriver à E_k (à partir de E_h) à la 1^{e}, à la 2^{e}, ..., à la $n^{\text{ième}}$ épreuve. Ce qui distingue du cas de Poisson, c'est la restriction : à partir de E_h.

On se rapproche du cas de Poisson quand on passe à la limite dans le cas régulier, en ce sens que la *fréquence moyenne de* E_k *à partir*

de E_h *dans* n *épreuves tend, dans le cas régulier, vers une limite* P_k *indépendante de l'état initial* E_h. Non seulement $\mathfrak{M} f_{hk}^{(n)} - P_k$ est infiniment petit avec $\frac{1}{n}$, mais nous savons même (p. 31) que cet infiniment petit est d'ordre au moins égal à celui de $\frac{1}{n}$ et nous avons appris à calculer la limite s_{hk} de $n[\mathfrak{M} f_{hk}^{(n)} - P_k]$. Observons bien que, si Π_{hk} et $P_{hk}^{(n)}$ tendent tous deux vers P_k, le second converge plus vite en général que le premier. Car on a vu qu'on peut majorer $\sum_n |P_{hk}^{(n)} - P_k|$ par une progression géométrique convergente, de sorte que $n(P_{hk}^{(n)} - P_k)$ tend vers zero, tandis que $n(\Pi_{hk}^{(n)} - P_k)$ tend vers une limite s_{hk} qui est en général $\neq 0$.

DISPERSION DANS LE CAS RÉGULIER

Dispersion d'une variable aléatoire. — Revenons maintenant au cas plus général d'une variable aléatoire quelconque. Il est intéressant d'évaluer l'ordre de grandeur de la dispersion de $Y_h^{(n)}$ et de constater comment varie cet ordre de grandeur avec h et surtout avec n. On peut, dans ce but, calculer différentes sortes d'écarts. Nous nous placerons encore ici dans le cas régulier pour voir ce que deviennent ces écarts quand n croît.

Soit d'abord $\mu_h^{(n)}$ l'écart quadratique moyen de $Y_h^{(n)}$. On a

$$[\mu_h^{(n)}]^2 = \mathfrak{M}[Y_h^{(n)} - \mathfrak{M} Y_h^{(n)}]^2 = \sum_j P_{hj}^{(n)} \left[x_j - \sum_k P_{hk}^{(n)} x_k \right]^2.$$

Dans le cas régulier, $\mu_h^{(n)}$ va tendre vers une limite μ, indépendante de h, donnée par la formule

$$(44) \qquad \mu^2 = \sum_j P_j [x_j - M]^2 \qquad \text{avec} \qquad M = \sum_k P_k x_k.$$

(μ est donc égal à l'écart quadratique moyen de la variable aléatoire Z définie page 68.)

On en déduit la valeur $\lambda_h^{(n)}$ de l'écart quadratique moyen de $X_h^{(n)}$ avec M; on aura

$$(\lambda_h^{(n)})^2 = (\mu_h^{(n)})^2 + (\mathfrak{M} Y_h^{(n)} - M)^2.$$

et, par suite, lorsque n croît indéfiniment, $\lambda_h^{(n)}$ (toujours au moins égal à $\mu_h^{(n)}$) tend aussi vers μ.

Application au mélange des urnes. — Calculons la limite μ quand n croît de l'écart quadratique moyen du nombre k de boules blanches dans l'urne U après n épreuves. On a, d'après les formules (10) de la page 51,

$$\mu^2 = \sum_k \mathrm{P}_k (k - \bar{k})^2 = \sum_k k^2 \mathrm{P}_k - (\bar{k})^2 = \sum_k k^2 \left(\frac{\mathrm{C}_u^k \mathrm{C}_v^{\mathrm{B}-k}}{\mathrm{C}_\mathrm{T}^\mathrm{B}} \right) - (\bar{k})^2.$$

D'où

$$\mu^2 = \sum_k k(k-1) \frac{\mathrm{C}_u^k \mathrm{C}_v^{\mathrm{B}-k}}{\mathrm{C}_\mathrm{T}^\mathrm{B}} + \sum_k k \frac{\mathrm{C}_u^k \mathrm{C}_v^{\mathrm{B}-k}}{\mathrm{C}_\mathrm{T}^\mathrm{B}} - \bar{k}^2$$

$$= \sum_k \frac{\mathrm{C}_{u-2}^{k-2} \mathrm{C}_v^{\mathrm{B}-k}}{\mathrm{C}_\mathrm{T}^\mathrm{B}} u(u-1) + \bar{k} - (\bar{k})^2$$

Or, par un raisonnement analogue à celui de la page 70, on voit que

$$\sum_k \mathrm{C}_{u-2}^{k-2} \mathrm{C}_v^{\mathrm{B}-k} = \mathrm{C}_{u+v-2}^{\mathrm{B}-2}$$

D'où

$$\mu^2 = u(u-1) \frac{\mathrm{C}_{u+v-2}^{\mathrm{B}-2}}{\mathrm{C}_\mathrm{T}^\mathrm{B}} + \bar{k}(1 - \bar{k}) \qquad \text{avec} \quad \bar{k} = \frac{\mathrm{B}u}{\mathrm{T}},$$

c'est-à-dire

$$\mu^2 = \frac{u(u-1)\mathrm{B}(\mathrm{B}-1)}{\mathrm{T}(\mathrm{T}-1)} + \frac{\mathrm{B}u}{\mathrm{T}} \left(1 - \frac{\mathrm{B}u}{\mathrm{T}} \right)$$

$$= \frac{\mathrm{B}u}{\mathrm{T}} \left\{ \frac{(u-1)(\mathrm{B}-1)}{\mathrm{T}-1} + 1 - \frac{\mathrm{B}u}{\mathrm{T}} \right\} = \frac{\mathrm{BN}uv}{\mathrm{T}^2(\mathrm{T}-1)}.$$

Ainsi

$$\mu^2 = \frac{\mathrm{BN}uv}{\mathrm{T}^2(\mathrm{T}-1)}. \tag{45}$$

On observe qu on a nécessairement

$$\mathrm{BN} \leqq \frac{\mathrm{T}^2}{4} \qquad \text{et} \qquad uv \leqq \frac{\mathrm{T}^2}{4},$$

d'où

$$\mu^2 \leqq \frac{\mathrm{T}^2}{16(\mathrm{T}-1)},$$

et la valeur maximum $\frac{T^2}{16(T-1)}$ de μ^2 pour T donné est d'ailleurs atteinte dans le cas de Laplace où $B = N = u = v$. Quand T est grand, on peut écrire approximativement

$$\mu \sim \sqrt{\frac{BNuv}{T^3}},$$

et le second membre est au plus égal à $\frac{\sqrt{T}}{4}$.

Dispersion de la moyenne arithmétique de la valeur observee au cours de n épreuves. — Revenons au cas général.

Markoff a aussi étudié le comportement de l'écart quadratique moyen $\rho_h^{(n)}$ de la moyenne arithmétique, $A_h^{(n)}$, des n premiers $Y_h^{(t)}$, $(\rho_h^{(n)})^2$ étant ainsi la valeur moyenne de $[A_h^{(n)} - M_h^{(n)}]^2$, c'est-à-dire de

$$\left\{\frac{(Y_h^{(1)} - \mathfrak{M}Y_h^{(1)}) + \ldots + (Y_h^{(n)} - \mathfrak{M}Y_h^{(n)})}{n}\right\}^2.$$

Si les $Y_h^{(t)}$ étaient indépendants, on sait que $n^2[\rho_h^{(n)}]^2$ serait égal à

$$(\mu_h^{(1)})^2 + \ldots + (\mu_h^{(n)})^2.$$

Comme $\mu_h^{(n)} \to \mu$,

$$\frac{(\mu_h^{(1)})^2 + \ldots + (\mu_h^{(n)})^2}{n}$$

tendrait vers μ^2 et par suite $\sqrt{n}(\rho_h^{(n)})$ vers μ, c'est-à-dire que $\rho_h^{(n)}$ serait un infiniment petit équivalent à $\frac{\mu}{\sqrt{n}}$.

En réalité, non seulement ce raisonnement n'est plus applicable puisque la valeur des $Y_h^{(n)}$ influe sur $Y_h^{(n+1)}$, mais encore le résultat du calcul ne subsiste pas entièrement.

Pour le voir, on va d'abord simplifier la marche du calcul, en remplaçant $M_h^{(n)}$ par M au moyen de l'égalité

$$(46)\quad n[\rho_h^{(n)}]^2 = n\mathfrak{M}[A_h^{(n)} - M_h^{(n)}]^2 = n\mathfrak{M}(A_h^{(n)} - M)^2 - n(M_h^{(n)} - M)^2.$$

Or, on a vu page 72 que, dans le cas régulier, $n^2[M_h^{(n)} - M]^2$ a une limite finie $(S_h)^2$. Dès lors $n[M_h^{(n)} - M]^2$ tend vers zéro et il reste à étudier la quantité

$$(47)\quad n[\delta_h^{(n)}]^2 = n\mathfrak{M}(A_h^{(n)} - M)^2$$

où $\delta_h^{(n)}$ a d'ailleurs une signification simple : c'est l'écart quadratique de $A_h^{(n)}$ avec M. On a

$$(48) \qquad n[\delta_h^{(n)}]^2 = \frac{1}{n}\mathfrak{M}\left\{\sum_{u=1}^{u=n} Y_h^{(u)} - M\right\}^2$$

$$= \frac{1}{n}\sum_{u=1}^{u=n}[\lambda_h^{(u)}]^2 + \frac{2}{n}\sum_{u\neq v}\mathfrak{M}\{(Y_h^{(u)} - M)(Y_h^{(v)} - M)\}$$

Or, $\lambda_h^{(u)}$ tend vers μ quand n croît indéfiniment, donc

$$\frac{1}{n}\sum_{u=1}^{u=n}(\lambda_h^{(u)})^2$$

tend vers μ^2. Il reste à étudier le comportement de la quantité

$$(48\ bis) \qquad L_n = \frac{2}{n}\sum_{u\neq v}\mathfrak{M}\{(Y_h^{(u)} - M)(Y_h^{(v)} - M)\}$$

$$= \frac{2}{n}\sum_{u\neq v}\left[\sum_i\sum_j(x_i - M)(x_j - M)K_{hij}^{(u,v)}\right]$$

$$= 2\sum_i\sum_j G_{i,j}^{(n)}(x_i - M)(x_j - M),$$

avec

$$(49) \qquad G_{i,j}^{(n)} = \frac{1}{n}\sum_{u\neq v}K_{hij}^{(u,v)},$$

où $K_{hij}^{(u,v)}$ désigne la probabilité que le Système, partant de l'état E_h, parvienne à l'état E_i à la $u^{\text{ième}}$ opération et à l'état E_j à la $v^{\text{ième}}$ opération.

Comme chaque couple u, v ne doit entrer qu'une fois dans la somme et sous la condition $u \neq v$, on peut supposer $u < v$. Alors

$$K_{hij}^{(u,v)} = P_{hi}^{(u)} P_{ij}^{(v-u)}$$

De sorte que

$$(50) \qquad nG_{i,j}^{(n)} = \sum_{1\leqq u<v\leqq n} P_{hi}^{(u)} P_{ij}^{(v-u)} = \sum_{v=2}^{v=n}\left[\sum_{u=1}^{u=v-1} P_{hi}^{(u)} P_{ij}^{(v-u)}\right]$$

$$= P_{hi}^{(1)}P_{ij}^{(1)} + \ldots + [P_{hi}^{(1)}P_{ij}^{(n-1)} + \ldots + P_{hi}^{(n-1)}P_{ij}^{(1)}].$$

Pour faciliter l'étude de la convergence, nous allons substituer

à $G_{ij}^{(n)}$, la quantité $S_{ij}^{(n)}$ définie par

$$n S_{ij}^{(n)} = P_{hi}^{(1)} \varphi_{ij}^{(1)} + [P_{hi}^{(1)} \varphi_{ij}^{(2)} + P_{hi}^{(2)} \varphi_{ij}^{(1)}] + \dots$$
$$+ [P_{hi}^{(1)} \varphi_{ij}^{(n-1)} + \dots + P_{hi}^{(n-1)} \varphi_{ij}^{(1)}]$$

où

$$(51) \qquad \varphi_{ij}^{(n)} = P_{ij}^{(n)} - P_j$$

La substitution ne modifie pas L_n, car elle revient à retrancher de L_n la quantité

$$\frac{2}{n} \sum_{ij} (x_i - M)(x_j - M) \left[\sum_{u<v} P_{hi}^{(u)} P_j \right]$$
$$= \frac{2}{n} \left[\sum_{i} \sum_{u<v} P_{hi}^{(u)} (x_i - M) \right] \left[\sum_{j} P_j (x_j - M) \right]$$

Or, le dernier crochet est nul.

D'autre part, $S_{ij}^{(n)}$ est le produit par $\frac{n-1}{n}$ de la moyenne arithmétique des $n-1$ premiers termes de la suite $\alpha_1, \dots, \alpha_n, \dots$ où

$$\alpha_n = P_{hi}^{(1)} \varphi_{ij}^{(n)} + \dots + P_{hi}^{(n)} \varphi_{ij}^{(1)}.$$

Si nous démontrons que α_n a une limite α, il en résultera que $S_{ij}^{(n)} \to \alpha$.

Or, si nous nous plaçons dans le cas régulier, nous savons que $P_{hi}^{(n)}$ tend vers P_i quand n croît indéfiniment et que les différences $\varphi_{ij}^{(n)} = P_{ij}^{(n)} - P_j$ forment une série absolument convergente, ces différences étant, en effet, en valeurs absolues inférieures aux termes d'une progression géométrique convergente (*voir* p. 31). Dans ces conditions, d'après le premier Lemme de la Note C, page 280, α_n tend nécessairement vers $P_i \sum_n \varphi_{ij}^{(n)}$.

Dès lors, on voit qu'on a

$$\lim_{n \to \infty} S_{ij}^{(n)} = P_i \sum_{n=1}^{n=\infty} \varphi_{ij}^{(n)} = P_i s_{ij},$$

$$\lim_{n \to \infty} L_n = 2 \sum_{i} \sum_{j} P_i s_{ij} (x_j - M)(x_i - M) = L \quad (^1),$$

---

(1) D'après la façon dont a été calculée la seconde somme du second membre,

où s_{ij} est la somme de la série convergente

$$s_{ij} = \sum_{n=1}^{n=\infty} [P_{ij}^{(n)} - P_j]$$

Et, finalement, on voit qu'on a

$$\lim_{n \to \infty} n [\delta_h^{(n)}]^2 = \mu^2 + L$$

Le premier membre étant certainement $\geqq 0$, nous pouvons poser $\sigma^2 = \mu^2 + L$, c'est-à-dire

$$(52) \qquad \sigma^2 = \sum_j P_j (x_j - M)^2 + 2 \sum_{ij} P_i s_{ij} (x_i - M)(x_j - M)$$

D'où, finalement,

$$\lim_{n \to \infty} n [\rho_h^{(n)}]^2 = \sigma^2,$$

quantité indépendante de l'état initial E_h. On a vu plus haut comment calculer directement les P_i et s_{ij} sans avoir à effectuer une suite d'itérations. On obtiendra ainsi directement σ^2.

Ainsi, nous avons trouvé que, dans le cas régulier, non seulement l'écart quadratique moyen de $A_h^{(n)}$, soit

$$\rho_h^{(n)} = \sqrt{\mathfrak{M} \left[\frac{Y_h^{(1)} + \ldots + Y_h^{(n)}}{n} - M_h^{(n)} \right]^2},$$

tend vers zéro avec $\frac{1}{n}$, mais si σ est $\neq 0$, *cet écart est un infiniment petit équivalent à* $\frac{\sigma}{\sqrt{n}}$, σ étant déterminé par la formule (52) où, d'ailleurs, il y a lieu de noter que *σ est indépendant de h*. On voit que la dépendance des $Y_h^{(n)}$ entre eux ne modifie pas l'ordre de l'infiniment petit $\rho_h^{(n)}$ qui reste celui de $\frac{1}{\sqrt{n}}$, mais modifie la valeur de la

---

elle est de la forme

$$L = 2 \sum_i P_i s_{ii} (x_i - M)^2 + 2 \sum_{i \neq j}{}' (P_i s_{ij} + P_j s_{ji})(x_i - M)(x_j - M),$$

où, dans Σ', chaque couple i, j avec $i \neq j$ ne doit être considéré qu'une fois.

partie principale qui est $\frac{\sigma}{\sqrt{n}}$ et non pas $\frac{\mu}{\sqrt{n}}$. Et, en général, σ sera différent de μ. On vérifie d'ailleurs que σ se réduit bien à μ dans le cas où les $Y_h^{(n)}$ sont indépendants et plus précisément lorsque la probabilité p_{ij} de passer de l'état E_i à l'état E_j est indépendante de l'état initial E_i. Dans ce cas,

$$p_{ij} = P_j = P_{ij}^{(n)},$$

et, par suite, $s_{ij} = 0$ et $\sigma = \mu$.

C'est Markoff qui a, le premier, mis en lumière les importants résultats découlant de l'étude précédente. I, les valeurs moyennes de $Y_h^{(n)}$ et de $A_h^{(n)} = \frac{Y_h^{(1)} + \ldots + Y_h^{(n)}}{n}$ ont la même limite dans le cas actuel comme dans le cas des événements indépendants, II, au contraire les limites des écarts quadratiques moyens de $Y_h^{(n)}$ et de $A_h^{(n)}\sqrt{n}$, qui seraient égales dans le cas de l'indépendance, sont en général distinctes (quand elles existent) dans le cas d'événements en chaîne.

Markoff avait montré que $\rho_h^{(n)}\sqrt{n}$ a une limite dans le cas positivement régulier. Il avait, en outre, trouvé une expression d'ailleurs assez compliquée de cette limite. La preuve de l'existence de cette limite dans le cas régulier le plus général ainsi que l'expression fournie par (52) de cette limite ont été obtenues indépendamment et au moyen de méthodes distinctes par M. Potoček [2] et par M. Fréchet [3, p. 31].

Nous indiquerons plus loin une méthode plus générale qui permet de calculer aussi les moments des divers ordres de $A_h^{(n)}$.

Remarque. — Quand σ n'est pas nul, l'écart quadratique moyen de $(Y_h^{(1)} - M) + \ldots + (Y_h^{(n)} - M)$ a pour partie principale $\sigma\sqrt{n}$ et croît indéfiniment. Quand σ est nul, le raisonnement tombe en défaut. Nous verrons plus loin (p. 153) qu'alors, dans le cas régulier, cet écart moyen reste borné et même converge vers une limite déterminée, de sorte que $\rho_h^{(n)}$ est d'ordre au moins égal à celui de $\frac{1}{n}$. Mais, il y a lieu d'abord de voir si σ peut s'annuler.

Il y a deux cas évidents où il en est ainsi : d'abord lorsque les épreuves considérées ne sont en aucune façon fortuites, c'est-à-dire que le système étant à une épreuve à l'état E_i, passe nécessairement (ou même simplement avec une probabilité égale à 1) à un état déterminé E_k. De sorte que pour chaque i, la suite des p_{ik} comporte un

terme égal à 1 et les autres nuls. Alors $A_h^{(n)} = \frac{1}{n}(Y_h^{(1)} + \dots + Y_h^{(n)})$ est une fonction certaine de h et n, son écart quadratique moyen avec M_h est nul et par suite $\rho_h^{(n)} = 0$, donc $n[\delta_h^{(n)}]^2$ tend vers zéro, $\sigma = 0$.

Dans un autre cas, le hasard joue sur les états successifs réalisés, mais n'a pas d'influence sur $A_h^{(n)}$, c'est celui où les valeurs possibles de X sont égales : $x_1 = x_2 = \dots = a$. Alors les p_{ik} pouvant être quelconques, $A_h^{(n)}$ reste un nombre certain (égal à a) et, là encore, σ est nul.

MM. Dœblin [2] et Mihoc [2] ont montré que σ est nul dans des cas plus généraux, mais où l'effet du hasard reste faible, cas qui seront traités maintenant.

Détermination générale des cas où $\sigma = 0$. — I. *Procédé direct.* — On a, dans le cas régulier,

$$\sigma^2 = \sum_i P_i z_i^2 + 2\sum_i \sum_j P_i s_{ij} z_i z_j,$$

avec $\sum_i P_i z_i = 0$. Puisque $z_i = x_i - M$, et en vertu des formules (9) et (9 *bis*) des pages 46, 48 $\left(\sum_j s_{ij} = 0 \text{ et } \sum_i P_i s_{ij} = 0\right)$, on a

$$\sigma^2 = \sum_i P_i x_i^2 - \left(\sum_i P_i x_i\right)^2 + 2\sum_i \sum_j P_i s_{ij} x_i x_j.$$

Si l'on avait remplacé les x_i par d'autres valeurs ξ_i, on aurait eu de même

$$f(\xi_1, \dots, \xi_r) = \sum_i P_i \xi_i^2 - \left(\sum_i P_i \xi_i\right)^2 + 2\sum_i \sum_j P_i s_{ij} \xi_i \xi_j \geqq 0.$$

Dire que σ est nul, c'est donc dire que la forme quadratique $f(\xi_1, \dots, \xi_r)$, qui est $\geqq 0$ quels que soient $\xi_1, \dots, \xi_r$, atteint un minimum pour $\xi_1 = x_1, \dots, \xi_r = x_r$. Dès lors, on aura

$$\frac{\partial f}{\partial \xi_1} = \dots = \frac{\partial f}{\partial \xi_r} = 0 \qquad \text{pour} \quad \xi_1 = x_1, \dots, \xi_r = x_r.$$

Et réciproquement, si les relations

$$\frac{\partial f}{\partial \xi_1} = 0, \quad \dots, \quad \frac{\partial f}{\partial \xi_r} = 0 \tag{53}$$

ont un système de solutions $x_1, \ldots, x_r$, alors en vertu de l'identité

$$2f = \sum_i \xi_i \frac{\partial f}{\partial \xi_i},$$

σ sera nul. Ainsi la condition nécessaire et suffisante pour que σ soit nul, c'est que $x_1, \ldots x_r$ vérifient les r équations homogenes (53).

Or,

$$\frac{1}{2}\frac{\partial f(x_1, \ldots, x_r)}{\partial x_k} = P_k x_k - M P_k + \sum_j P_k s_{kj} x_j + \sum_i P_i s_{ik} x_i$$

$$= P_k(x_k - M) + P_k \sum_j s_{kj}(x_j - M) + \sum_i P_i s_{ik}(x_i - M)$$

Dès lors, il doit y avoir un système de solutions en $x_1, \ldots, x_r$ et M des équations

$$\sum_i P_i(x_i - M) = 0,$$

$$P_k(x_k - M) + P_k \sum_j s_{kj}(x_j - M) + \sum_i P_i s_{ik}(x_i - M) = 0,$$

ou encore un système de solutions en $z_1, \ldots, z_r$ du système

$$\sum_i P_i z_i = 0,$$

$$P_k z_k + P_k \sum_j s_{kj} z_j + \sum_i P_i s_{ik} z_i = 0 \qquad (k = 1, \ldots, r).$$

D'ailleurs, en ajoutant les r dernières équations de ce système, on obtient $\sum_k P_k z_k = 0$, en vertu des relations $\sum_k P_k s_{kj} = 0$ et $\sum_k s_{ik} = 0$. Dès lors, la première équation du système est une conséquence des autres. En résumé, *dans le cas régulier, pour que σ soit nul, il faut et il suffit que les r équations linéaires et homogènes à r inconnues $z_1, \ldots, z_r$:*

$$(54) \qquad P_k z_k + P_k \sum_j s_{kj} z_j + \sum_j P_j s_{jk} z_j = 0 \qquad (k = 1, \ldots, r),$$

où $z_j = x_j - \sum_i P_i x_i$, soient vérifiées.

Inversement, quand les p_{ik} (et par suite aussi les P_i et les s_{ik}) sont donnés, ce système admet par rapport à des inconnues arbitraires $z_1, \ldots, z_r$ au moins un système de solutions $t_1, \ldots, t_r$. Alors, prenant pour M une constante arbitraire, il suffira de prendre

$$x_j = t_j + M \qquad \text{d'où} \qquad \sum_j P_j x_j = \sum_j P_j t_j + M = M,$$

pour que les p_{ik} donnés et les x_j ainsi déterminés correspondent à une valeur de σ nulle. De plus, on aura ainsi la méthode la plus générale, toujours dans le cas régulier, pour déterminer les x_j de sorte que $\sigma = 0$.

Un premier système de solutions de (54) se présente, quelles que soient les quantités p_{ik}, à savoir $z_1 = \ldots = z_r = 0$.

C'est-à-dire que σ est nécessairement nul quand les valeurs données $x_1, \ldots, x_r$ sont toutes égales entre elles (et par suite égales à M).

Mais σ peut être nul dans d'autres cas. Il suffit que les p_{ik} soient tels que le système des équations (54) admette un système de solutions non toutes nulles, ce qui a lieu quand est nul le déterminant

$$\delta = \begin{vmatrix} P_1(1+s_{11}) + P_1 s_{11} & P_2 s_{21} + P_1 s_{12} & \cdots & P_r s_{r1} + P_1 s_{1r} \\ P_1 s_{12} + P_2 s_{21} & P_2(1+s_{22}) + P_2 s_{22} & \cdots & P_r s_{r2} + P_2 s_{2r} \\ & P_2 s_{23} + P_3 s_{32} & & \\ \cdots & \cdots & \cdots & \cdots \\ P_1 s_{1r} + P_r s_{r1} & P_2 s_{2r} + P_r s_{r2} & \cdots & P_r(1+s_{rr}) + P_r s_{rr} \end{vmatrix}$$

Nous savons alors que si ce déterminant est nul, il y aura au moins un système de solutions en z, non toutes nulles, c'est-à-dire au moins un système de valeurs non toutes égales des x_i, pour lesquelles σ est nul.

Comme on sait (p. 43 et 46) calculer, dans le cas régulier, les P et les s en fonction des p_{ik}, il serait intéressant d'exprimer aussi ce déterminant en fonction des p_{ik}. Nous verrons plus loin (p. 86) qu'une fois ainsi exprimé, ce déterminant ne reste pas identiquement nul quand les p_{ik} varient. Mais nous nous contenterons d'insister sur la conséquence évidente de la méthode précédente : pour que σ soit nul sans que les x_i soient égaux, il faut d'abord que les p_{ik} satisfassent d'abord à une certaine relation $\delta = 0$ et qu'ensuite les x_i soient choisis convenablement. Ce choix des x_i qui ne s'effectue qu'à une constante addi-

tive près peut se réaliser par la résolution d'un système de relations linéaires (54).

II. *Procédé basé sur une décomposition en carrés de* σ^2. — On peut obtenir des conditions de formes plus simples en décomposant σ^2 en carrés. A cet effet, écrivons l'expression de σ^2, en introduisant le signe $\theta_i = \sum_j s_{ij} z_j$. On a

$$\sigma^2 = \sum_i P_i z_i^2 + 2\sum_i P_i z_i \theta_i = \sum_i P_i (z_i + \theta_i)^2 - \sum_i P_i \theta_i^2$$

$$= \sum_i \left(\sum_k P_k p_{ki} \right) (z_i + \theta_i)^2 - \sum_k P_k \theta_k^2$$

$$= \sum_k P_k \left\{ \sum_i p_{ki} (z_i + \theta_i)^2 - \theta_k^2 \right\}$$

Or, on a, d'après la formule (8 *ter*) de la page 47,

$$\sum_i p_{ki}(z_i + \theta_i) = \sum_j \left\{ p_{kj} + \sum_i p_{ki} s_{ij} \right\} z_j = \sum_j \{ s_{kj} + P_j \} z_j = \theta_k.$$

de sorte que, dans σ^2,

$$\sum_i p_{ki}(z_i + \theta_i)^2 = \sum_i p_{ki} [(z_i + \theta_i - \theta_k) + \theta_k]^2$$

$$= \sum_i p_{ki}(z_i + \theta_i - \theta_k)^2 + \theta_k^2.$$

D'où, enfin, la décomposition annoncée

$$\sigma^2 = \sum_k P_k G_k,$$

avec

$$G_k = \sum_i p_{ki}(z_i + \theta_i - \theta_k)^2.$$

Dès lors : *dans le cas régulier le plus général, la condition nécessaire et suffisante pour que l'écart quadratique moyen de* $Y_h^{(1)} + \ldots + Y_h^{(n)} - nM$ *reste borné quand n croît est que, pour tout*

couple i, k, l'on ait

$$(55)\qquad P_k p_{ki}(z_i + \theta_i - \theta_k) = 0 \qquad \text{avec} \quad \theta_i = \sum_j s_{ij} z_j$$

Considérons d'abord le cas simple où tous les p_{ik} sont $\neq 0$. Alors on est dans le cas positivement régulier, les P_k sont $\neq 0$, donc les G_k sont nuls et l'on doit avoir pour tout couple i, k

$$z_i + \theta_i - \theta_k = 0$$

En particulier, on voit, pour $k = i$, qu'on devra avoir $z_i = 0$. Ainsi, *dans le cas où les p_{ik} sont tous* $\neq 0$ (et plus généralement, dans le cas positivement régulier, quand la diagonale principale formée des p_{ii} n'a aucun terme nul), *le seul cas où σ soit nul est celui où les x_i sont égaux entre eux.*

D'après cela, on voit bien que si l'on exprime δ en fonction des p_{ik}, δ n'est pas identiquement nul, puisqu'il est nécessairement $\neq 0$ si les p_{ik} sont tous $\neq 0$.

Nous avons noté précédemment un exemple où cependant les p_{ik} sont tels que σ^2 soit nul quelles que soient les valeurs des x_i. Mais cet exemple concernait un cas singulier, nous venons de voir qu'il n'aurait pu se présenter si les p_{ik} avaient été tous $\neq 0$.

Et même, comme l'a signalé M. Mihoc [2], on peut montrer qu'il est impossible de rencontrer un tel exemple en dehors du cas singulier. Supposons, en effet, $\sigma \equiv 0$ pour des p_{ki} determinés, quels que soient les x_i. Pour tout couple k, i, on aura $P_k p_{ki} = 0$ ou

$$z_i + \sum_j (s_{ij} - s_{kj}) z_j = 0,$$

c'est-à-dire

$$x_i + \sum_j (s_{ij} - s_{kj}) x_j = \sum_j P_j x_j,$$

quels que soient les x_j. D'où

$$\begin{cases} s_{ij} - s_{kj} = P_j & \text{pour} \quad j \neq i. \\ s_{ii} - s_{ki} + 1 = P_i. \end{cases}$$

Et en substituant dans la relation connue

$$P_i - p_{ki} = \sum_j s_{kj} p_{ji} - s_{ki},$$

on aura

$$P_i - p_{ki} = \sum_j (s_{ij} - P_j) p_{ji} + p_{ii} - (s_{ii} + 1 - P_i),$$

et comme la sommation donne $s_{ii} - p_{ii} + P_i - P_i$, on aura finalement

$$p_{ki} = 1$$

Ainsi, pour chaque valeur de k, ou bien $P_k = 0$, ou bien on a, pour cette valeur de k, et pour chaque valeur de i, $p_{ki} = (0 \text{ ou } 1)$.

Désignons par k' les indices des états tels que $P_{k'} \neq 0$ et par k'' les autres. $\left(\text{Il y a un } k' \text{ au moins en vertu de } \sum_k P_k = 1\right)$. S'il existe au moins un k'' on aura

$$0 = P_{k''} = \sum_j P_j p_{jk''} = \sum_{k'} P_{k'} p_{k'k''},$$

donc tous les $p_{k'k''}$ sont nuls. Cette remarque sera généralisée (p. 167) On vient de voir que pour tout k', les $p_{k'i}$ sont égaux à 0 ou 1 et les $p_{k'k''}$ sont nuls, s'il en existe. Comme $\sum_i p_{k'i} = 1$, on voit que pour tout k'_1 des k', il existe un autre élément k'_2 des k' et un seul tel que $p_{k'_1 k'_2} = 1$ et par suite $p_{k_1 i} = 0$ pour $i \neq k'_2$. On peut alors diviser les états k' en un certain nombre de cycles tels que $E_{k'_1}$, $E_{k'_2}$, ..., $E_{k'_\nu}$, E_{k_1} tels que le Système matériel entré dans un cycle y reste presque sûrement et le parcourt dans un sens déterminé. Alors on ne peut être dans le cas régulier. Il est clair, en effet, et l'on vérifie facilement que $P_{k_1 i}^{(1+m\nu)} = 1$ si $i = k'_2$ et $= 0$ dans le cas contraire, de sorte que $P_{k'_1 k'_2}^{(n)}$ prend périodiquement les valeurs 1 et 0 au lieu de converger. Même si chaque cycle ne comprend qu'un seul terme, alors $P_{kk_1}^{(n)}$ pourrait bien avoir une limite, mais non une limite indépendante de k.

III. *Une condition plus simple dans le cas positivement régulier.* — On doit à MM. Doeblin [1, 2] et Mihoc [2], qui l'ont obtenu à peu près simultanément, une condition plus intuitive que les conditions (54) pour que σ soit nul dans le cas positivement régulier (que les p_{ii} soient ou non $\neq 0$).

Nous l'obtiendrons d'ailleurs ici d'une manière différente de celles

de ces auteurs, comme conséquence des conditions (55), qui ont été déduites d'une étude directe de σ^2. Pour tout couple k, i, tel que p_{ki} soit $\neq 0$, on aura

$$z_i + \theta_i - \theta_k = 0.$$

Considérons maintenant une permutation circulaire de valeurs de i : $i_1, i_2, \ldots, i_\alpha, i_1$. Appelons *cycle* d'états $E_{i_1}, E_{i_2}, \ldots, E_{i_\alpha}, E_{i_1}$, une permutation circulaire d'états telle qu'une suite d'épreuves faisant parcourir au Système considéré ce cycle d'états ne soit pas presque impossible, c'est-à-dire ait une probabilité positive, autrement dit que $p_{i_1 i_2}, p_{i_2 i_3}, \ldots, p_{i_{\alpha-1} i_\alpha}, p_{i_\alpha i_1}$ sont tous $\neq 0$.

Alors on aura

$$z_{i_2} + \theta_{i_2} - \theta_{i_1} = 0, \quad \ldots, \quad z_{i_\alpha} + \theta_{i_\alpha} - \theta_{i_{\alpha-1}} = 0, \quad z_{i_1} + \theta_{i_1} - \theta_{i_\alpha} = 0,$$

d'où

$$z_{i_1} + z_{i_2} + \ldots + z_{i_\alpha} = 0$$

ou

$$\frac{x_{i_1} + x_{i_2} + \ldots + x_{i_\alpha}}{\alpha} = M.$$

Réciproquement supposons que, pour tout cycle d'états, la moyenne arithmétique des valeurs x_i correspondante ait une valeur fixe A indépendante du cycle et quel que soit le nombre des états de ce cycle, distincts ou non. Et considérons la quantité

$$\mathcal{L}_{j,n}^{(\nu)} = \mathfrak{M}_{E_j}[(X^{(1)} - A) + \ldots + (X^{(n)} - A)]^\nu,$$

qui est égale à

$$\sum_{i_1, \ldots, i_n} p_{j i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}[(x_{i_1} - A) + \ldots + (x_{i_n} - A)]^\nu.$$

On peut, sans changer $\mathcal{L}_{j,n}^{(\nu)}$, remplacer la sommation complète Σ par la sommation incomplète Σ' où l'on convient de ne faire entrer que les termes contenant des produits $p_{j i_1} \ldots p_{i_{n-1} i_n}$ qui ne sont pas nuls. Considérons un terme de Σ'. Pour $n > r$, les nombres $i_1, \ldots, i_n$ ne peuvent être tous distincts; on a, par exemple, $i_\alpha = i_{\alpha+\beta}$, alors la suite d'états

$$E_{i_1}, \; E_{i_2}, \; \ldots, \; E_{i_\alpha}, \; \ldots, \; E_{i_{\alpha+\beta-1}}, \; E_{i_\alpha}, \; E_{i_{\alpha+\beta+1}}, \; \ldots, \; E_{i_n}$$

comprend un cycle $E_{i_\alpha} \ldots E_{i_{\alpha+\beta}}$. Pour celui-ci, on a, en posant

$t_i = x_i - A$,

$$t_{i_{\sigma+1}} + t_{i_{\sigma+2}} + \dots + t_{i_{\sigma+\beta}} = 0$$

Considérons la suite restante

$$E_{i_1}, \ E_{i_2}, \ \dots, \ E_{i_\alpha}, \ E_{i_{\sigma+\beta+1}}, \ \dots, \ E_{i_n}$$

correspondant aux autres valeurs des t_{i_k}. Si elle comprend plus de r termes, elle comprend encore un cycle, et l'on ne change pas $t_{i_1} + \dots + t_{i_n}$ en supprimant les valeurs de t correspondantes. Et ainsi de suite. Toute somme $t_{i_1} + \dots + t_{i_n}$ de Σ' est donc égale à une somme de même forme, mais ne comprenant pas plus de r termes. Chacun de ces termes ne pouvant avoir que les valeurs $x_1 - A, \dots, x_r - A$, nous voyons que, dans Σ', tout terme $t_{i_1} + \dots + t_{i_n}$ ne peut avoir qu'un nombre de valeurs non seulement fini mais borné quand n croît. Pour n déterminé, la puissance $\nu^{\text{ième}}$ de chacune de ces valeurs sera multipliée dans Σ' par une quantité au plus égale à la probabilité que $[(x_{i_1} - A) + \dots + (x_{i_n} - A)]$ prenne cette valeur. Dès lors, $\mathcal{L}^{(\nu)}_{j,n}$ reste borné quand n croît. Prenons $\nu = 1$.

En posant

$$\mathfrak{N}^{(\nu)}_{j,n} = \mathfrak{M}_{E_j}[(X^{(1)} - M) + \dots + (X^{(n)} - M)]^\nu,$$

on a

$$\mathcal{L}^{(1)}_{j,n} = \mathfrak{N}^{(1)}_{j,n} + n(M - A).$$

Si nous nous plaçons dans le cas régulier, $\mathfrak{N}^{(1)}_{j,n}$ comme $\mathcal{L}^{(1)}_{j,n}$ restent bornés quand n croît; ceci exige que $M = A$, ce qui était à prévoir et se trouve démontré. Alors

$$\mathcal{L}^{(\nu)}_{j,n} = \mathfrak{N}^{(\nu)}_{j,n}$$

pour tout entier ν. Donc, pour toute valeur fixe de ν, les moments $\mathfrak{N}^{(\nu)}_{j,n}$ restent bornés quand n croît. Il en est ainsi en particulier pour $\nu = 2$. Or $\dfrac{\mathfrak{N}^{(2)}_{j,n}}{n} = n[\delta^{(n)}_j]^2$ qui tend vers σ^2. Donc $\sigma^2 = 0$.

En résumé, *dans le cas positivement régulier, la condition nécessaire et suffisante pour que* σ^2 *soit nul* (c'est-à-dire pour que l'écart moyen de $[Y^{(1)}_h - M + \dots + Y^{(n)}_h - M]$ reste borné quand n croît), *est que la moyenne arithmétique des valeurs* x_i *correspondant à un cycle d'états garde une valeur indépendante du nombre d'états de ce cycle et du cycle lui-même.*

Dispersion des fréquences. — Pour juger de la façon dont les fréquences se dispersent ou se concentrent, nous calculerons aussi leurs écarts quadratiques moyens. La valeur moyenne de $F_{hk}^{n,N}$ (p. 72), est, comme on l'a vu, P_{hk}^{n}, probabilité constante d'un des N groupes à l'autre. Et l'on sait qu'alors son écart quadratique moyen est

$$\mathcal{F}_{hk}^{n,N} = \sqrt{\frac{P_{hk}^{n}\left[1 - P_{hk}^{n}\right]}{N}}$$

D'où

$$\lim_{n \to \infty} \mathcal{F}_{hk}^{(n),N} = \sqrt{\frac{P_k(1 - P_k)}{N}}.$$

Passons maintenant à la dispersion de $f_{hk}^{(n)}$, c'est-à-dire de la fréquence avec laquelle dans un groupe de n épreuves on trouve l'état E_k (à quelque rang que ce soit dans ce groupe) à partir de l'état initial E_h. On a vu, page 73, que

$$\mathfrak{M} f_{hk}^{(n)} = \Pi_{hk}^{(n)} \qquad \text{et} \qquad \lim_{n \to \infty} \mathfrak{M} f_{hk}^{(n)} = P_k$$

Les limites de $\mathfrak{M} F_{hk}^{(n),N}$ et de $\mathfrak{M} f_{hk}^{(n)}$ sont les mêmes. Il n'en est pas ainsi pour leurs dispersions.

L'écart quadratique moyen $\mathcal{E}_{hk}^{n}$ de $f_{hk}^{(n)}$ s'obtient indirectement comme cas particulier. Si $Y_h^{(n)}$ est égal à 1 quand E_k se produit (n épreuves après E_h) et à zéro dans le cas contraire, alors

$$A_h^{(n)} = \frac{1}{n}\left[Y_h^{(1)} + \ldots + Y_h^{(n)}\right] = f_{hk}^{(n)},$$

et, par suite, $\rho_h^{(n)}$ se réduit à $\mathcal{E}_{hk}^{(n)}$, de même $\delta_h^{(n)}$ se réduit à l'écart quadratique moyen $\mathcal{G}_{hk}^{(n)}$ de $f_{hk}^{(n)}$ avec P_k. On a trouvé, page 79,

$$\lim_{n \to \infty} n\left[\rho_h^{(n)}\right]^2 = \sigma^2$$

avec

$$\sigma^2 = \sum_j P_j(x_j - M)^2 + 2\sum_{ij} P_i s_{ij}(x_i - M)(x_j - M)$$

$$= \sum_j P_j x_i^2 - M^2 + 2\sum_{ij} P_i s_{ij} x_i x_j.$$

Or, ici $x_j = 1$ si $j = k$, $x_j = 0$ si $j \neq k$ et $M = P_k$. On a donc

$$\lim_{n \to \infty} n\left[\mathcal{E}_{hk}^{(n)}\right]^2 = (\psi_k)^2$$

avec

$$(\psi_k)^2 = P_k(1 - P_k + 2s_{kk}).$$

En résumé, *on a dans le cas régulier*

$$\lim_{n \to \infty} n[\mathcal{G}_{hk}^{(n)}]^2 = \lim_{n \to \infty} n[\mathcal{E}_{hk}^{(n)}]^2 = P_k(1 - P_k + 2s_{kk}),$$

et cette limite est indépendante de l'état initial E_h.

Ainsi, non seulement $\mathcal{E}_{hk}^{(n)}$ est infiniment petit avec $\frac{1}{n}$, comme l'a prouvé M. von Mises [1, p. 554], mais encore comme l'avait prouvé antérieurement Markoff, $\mathcal{E}_{hk}^{(n)}$ est d'un ordre au moins égal à celui de $\frac{1}{\sqrt{n}}$.

En général, $\psi_k \neq 0$ et alors $\mathcal{E}_{hk}^{(n)}$, $\mathcal{G}_{hk}^{(n)}$ ont pour partie principale $\frac{\psi_k}{\sqrt{n}}$. Si $\psi_k = 0$, ce qui peut arriver, même dans le cas régulier (si par exemple, $p_{1k} = p_{2k} = \ldots = p_{rk} = 0$ et par suite $P_k = 0$), pour une valeur particulière de k, alors $\mathcal{E}_{hk}^{(n)}$ sera, soit, identiquement nul, soit d'un ordre supérieur à celui de $\frac{1}{\sqrt{n}}$ et même, plus précisément, d'un ordre au moins égal à celui de $\frac{1}{n}$, comme on le verra page 153 pour $\rho_h^{(n)}$ quand $\sigma = 0$.

Application au cas de deux états possibles. — En prenant $r = 2$, on peut poser $p_{12} = a$, $p_{21} = b$ et alors $p_{11} = 1 - a$, $p_{22} = 1 - b$. On doit avoir $0 \leqq {a \atop b} \leqq 1$.

Nous verrons plus loin, page 118, à quelles conditions doivent satisfaire a, b pour qu'on soit dans le cas régulier. Dès maintenant, nous savons qu'on sera dans ce cas, et même dans le cas positivement régulier, si $0 < {a \atop b} < 1$ (c'est-à-dire en général), puisque alors tous les p_{hk} sont $\neq 0$.

Si l'on est dans le cas régulier, il est facile de calculer les probabilités limites P_1, P_2, au moyen des équations ($\mathcal{E}$), page 44, qui deviennent ici

$$P_k = P_1 p_{1k} + P_2 p_{2k}, \qquad P_1 + P_2 = 1 \qquad (k = 1, 2);$$

d'où

$$P_1 = \frac{b}{a+b}, \qquad P_2 = \frac{a}{a+b}. \tag{56}$$

[Observons d'ailleurs que la division est possible, c'est-à-dire que $a + b \neq 0$. Sans quoi, on aurait $a = b = 0$, $p_{12} = p_{21} = 0$, c'est le cas où le changement d'état est impossible (ou du moins de probabilité nulle) en une épreuve, et par conséquent aussi en n épreuves. Donc $P_{11}^{(n)} = 1$, $P_{21}^{(n)} = 0$ et ces deux quantités ne pouvant avoir la même limite, on ne serait pas dans le cas régulier.] Pour qu'on soit dans le cas régulier, il faudra donc que a et b ne soient pas tous deux nuls

On a, de même, les quantités s_{hk} par les équations (8 *bis*, 9), page 46, qui deviennent ici :

$$s_{h1} - (s_{h1} p_{11} + s_{h2} p_{21}) = p_{h1} - P_1, \qquad s_{h2} = -s_{h1}, \qquad (h = 1, 2),$$

d'où

$$s_{11} = \frac{a}{a+b}\left(\frac{1}{a+b} - 1\right) = -s_{12},$$

$$s_{21} = \frac{b}{a+b}\left(1 - \frac{1}{a+b}\right) = -s_{22}$$

On sait que la fréquence $f_{hk}^{(n)}$ a pour valeur moyenne $\Pi_{hk}^{(n)}$ et que $\Pi_{hk}^{(n)}$ tend vers P_k. De plus, $n[\Pi_{hk}^{(n)} - P_k]$ tend vers s_{hk} que nous venons de calculer. Enfin les écarts quadratiques moyens $\mathcal{E}_{hk}^{(n)}$, $\mathcal{G}_{hk}^{(n)}$ de $f_{hk}^{(n)}$ avec sa valeur moyenne $\Pi_{hk}^{(n)}$ ou avec la limite P_k de celle-ci, tendent vers 0 avec $\frac{1}{n}$, et l'on a même

$$\lim_{n \to \infty} n[\mathcal{G}_{hk}^{(n)}]^2 = \lim_{n \to \infty} n[\mathcal{E}_{hk}^{(n)}]^2 = [\psi_k]^2$$

avec

$$(\psi_k)^2 = P_k[1 - P_k + 2 s_{kk}].$$

On peut supprimer l'indice k de ψ_k, car on a ici

$$(\psi_1)^2 = \frac{ab}{(a+b)^2}\left[\frac{2}{a+b} - 1\right] = (\psi_2)^2 = \psi^2. \tag{57}$$

Si, maintenant, on considère une variable aléatoire prenant les valeurs x_1 à l'état E_1, x_2 à l'état E_2, on a pour limite de la valeur moyenne qu'elle prend après n épreuves :

$$M = P_1 x_1 + P_2 x_2 = \frac{b x_1 + a x_2}{a+b}.$$

Si $A_h^{(n)}$ est la moyenne arithmétique des valeurs prises par cette valeur lors des n premières épreuves, $A_h^{(n)}$ a pour écart quadratique

moyen $\rho_h^{(n)}$ et $n\sqrt{\rho_h^{(n)}}$ tend vers σ avec

$$\sigma^2 = \mu^2 + L,$$

$$\begin{aligned}\mu^2 &= P_1(x_1 - M)^2 + P_2(x_2 - M)^2 \\ &= P_1 x_1^2 + P_2 x_2^2 - (P_1 x_1 + P_2 x_2)^2 = P_1 P_2 (x_1 - x_2)^2.\end{aligned}$$

$$\mu^2 = \frac{ab}{(a+b)^2}(x_1 - x_2)^2,$$

$$\begin{aligned}L &= 2\sum_i \sum_j P_i s_{ij}(x_i - M)(x_j - M) \\ &= 2P_1 s_{11}(x_1 - M)[(x_1 - M) - (x_2 - M)] + 2P_2 s_{21}(x_2 - M)[(x_1 - M) - (x_2 - M)] \\ &= 2(x_1 - x_2)\{P_1 s_{11}(x_1 - M) + P_2 s_{21}(x_1 - M)\} \\ &= \frac{2ab}{(a+b)^2}\left(\frac{1}{(a+b)} - 1\right)(x_1 - x_2)^2.\end{aligned}$$

Ainsi

$$(58) \qquad \sigma^2 = \frac{ab}{(a+b)^2}(x_1 - x_2)^2\left[\frac{2}{(a+b)} - 1\right] = \psi^2(x_1 - x_2)^2$$

Dans le cas où la probabilité d'arriver à E_k en une épreuve est indépendante de l'état initial E_h, c'est-à-dire si $p_{1k} = p_{2k}$, on a $1 - a = b$, $a = 1 - b$, c'est-à-dire $(a+b) = 1$ et réciproquement. On voit que dans ce cas on peut écrire

$$p_{1k} = p_{2k} = P_k \qquad \text{avec} \quad P_2 = a, \; P_1 = b = 1 - a$$

On est dans le cas de Bernoulli et l'écart quadratique moyen $\mathcal{E}_{hk}^{(n)}$ de la fréquence $f_{hk}^{(n)}$ a pour partie principale $\frac{\psi}{\sqrt{n}} = \sqrt{\frac{a(1-a)}{n}}$. Dans ce cas, $L = 0$ et

$$\mu = \sigma = \sqrt{a(1-a)}\,|x_1 - x_2|$$

En revenant au cas des chaînes, c'est-à-dire où $a + b \neq 1$, on voit que la formule (57) peut s'écrire

$$\psi^2 = P_1 P_2 \frac{1 + p_{11} - p_{21}}{1 - p_{11} + p_{21}}.$$

C'est la formule obtenue par Markoff [15, p. 559, en bas] avec d'autres notations et par Hostinský [16, p. 28], leurs méthodes étant différentes entre elles et différentes de la nôtre.

Markoff a donné une application intéressante de cette formule en prenant pour succession d'épreuves, la succession de la lecture des lettres d'un texte russe choisi au hasard et en désignant par E_1 la lecture d'une voyelle, par E_2 celle d'une consonne. Alors p_{12}, par exemple, sera la probabilité qu'une consonne suive immédiatement une voyelle. Si la fréquence des consonnes est indépendante de la nature des lettres qui les précèdent immédiatement, on aura $p_{12} = p_{22}$ et alors $p_{11} = p_{21}$. On peut obtenir par une statistique, les valeurs empiriques des p_{ik}. Que les valeurs obtenues pour $p_{12} - p_{22}$ et $p_{11} - p_{21}$ soient faibles ou non, on pourrait se demander si le fait qu'elles ne sont pas nulles n'est pas dû simplement à ce que les fréquences ne sont que des valeurs approchées des probabilités. Alors, en calculant la dispersion observée, d'abord comme si les consonnes et voyelles se succédaient indifféremment, puis dans l'hypothèse contraire et en comparant les résultats obtenus aux résultats fournis par les expressions de σ et de μ, on pourra se rendre compte quelle est l'hypothèse à adopter. Markoff a conclu de ses statistiques qu'on avait bien une dépendance entre les résultats de deux lectures successives

Retour au mélange des urnes — *Cas où le nombre des boules est grand* 1° Dans le cas simple de Laplace, où $u = v = B = N$, on a vu d'après la formule (11) que les P_k sont proportionnels aux carrés des probabilités $\varpi_k^{(u)} = C_u^k \frac{1}{2^u}$ des répétitions k d'un événement de probabilité $\frac{1}{2}$ dans u épreuves Cette remarque incite à considérer le cas où u est grand, cas où nous connaissons l'expression asymptotique

$$\varpi_k^{(u)} \sim \frac{e^{-y^2}}{\sqrt{\frac{\pi}{2} u}},$$

où

$$y = \frac{k - \frac{u}{2}}{\sqrt{\frac{u}{2}}}.$$

Cette expression est légitime quand y reste dans un intervalle fini indépendant de u, cas où nous allons nous placer pour simplifier, mais elle l'est même dans des cas plus généraux (*voir*, par exemple, Fréchet, [188, p. 92]). On a alors

$$C_u^k \sim 2^u \frac{e^{-y^2}}{\sqrt{\frac{\pi}{2} u}} \qquad \text{d'où} \quad P_k \sim \frac{2^{2u}}{C_{2u}^u \frac{\pi}{2} u} e^{-2y^2}.$$

En désignant par P_K le terme maximum, comme la valeur de y correspondante tend vers zéro avec u, on a

$$P_K \sim \frac{2^{2u}}{C_{2u}^u \frac{\pi}{2} u} \qquad \text{d'où} \quad P_k \sim P_K e^{-t^2},$$

en posant

$$t = y\sqrt{2},$$

c'est-à-dire

$$t = \frac{2}{\sqrt{u}}\left(k - \frac{u}{2}\right).$$

Ainsi, l'on voit que la loi de probabilité définie par l'expression de P_k en fonction de k, est asymptotiquement équivalente, quand le nombre de boules croît, à une « seconde loi de Laplace », où la valeur moyenne de k est $\frac{u}{2}$ et où l'écart quadratique moyen de k est $\frac{u}{2\sqrt{2}}$.

2° Dans le cas général où les u, v, B, N ne sont pas nécessairement égaux mais où les nombres k, $B-k$, $u-k$, $N-u+k$, $v = T-u$, et par suite aussi u, B, N et T sont tous très grands, on peut appliquer la formule asymptotique de Stirling $n! \sim n^n e^{-n}\sqrt{2\pi n}$ à l'expression

$$P_k = \frac{C_B^k C_N^{u-k}}{C_T^u} = \frac{B!\,N!}{k!(B-k)!(u-k)!(N-u+k)!}\,\frac{u!(T-u)!}{T!}.$$

On trouve

$$P_k \sim \frac{B^B N^N u^u v^v}{T^T}\,\frac{1}{k^k(B-k)^{B-k}(u-k)^{u-k}(N-u+k)^{N-u+k}}$$
$$\times \frac{1}{\sqrt{2\pi}}\sqrt{\frac{BNuv}{k(B-k)(u-k)(N-u+k)T}}.$$

Comme la limite de la valeur moyenne de $\frac{k}{u}$ quand n croît est $\frac{B}{T}$, on va, par raison d'analogie avec le cas 1°, poser

$$k - u\frac{B}{T} = x\rho,$$

où ρ est une quantité indépendante de k [1] et tendant vers l'infini avec T

(1) Par analogie avec ce qu'on fait dans le cas de Bernoulli, il serait naturel de prendre pour $\frac{\rho}{\sqrt{2}}$ l'écart quadratique moyen de k calculé plus haut, page 75,

mais de façon que $\frac{\rho}{u}$ tende vers zéro. Comme on a alors

$$(59)\quad \left\{\begin{aligned} k &= \frac{uB}{T} + x\rho, & B - k &= \frac{Bv}{T} - x\rho \\ u - k &= \frac{uN}{T} - x\rho, & N - u + k &= \frac{Nv}{T} + x\rho, \end{aligned}\right.$$

nous voyons que la quantité sous le dernier radical peut s'écrire

$$\frac{T^3}{BNuv}\,\frac{1}{\left(1+\frac{x\rho T}{uB}\right)\left(1-\frac{x\rho T}{Bv}\right)\left(1-\frac{x\rho T}{uN}\right)\left(1-\frac{x\rho T}{Nv}\right)}.$$

En ne cherchant à évaluer P_k que pour des valeurs de k pour lesquelles x reste borné quand T croît, l'expression asymptotique de la quantité ci-dessus sera $\frac{T^3}{BNuv}$ si nous supposons que les quantités

$$\frac{\rho T}{uB},\quad \frac{\rho T}{Bv},\quad \frac{\rho T}{uN},\quad \frac{\rho T}{Nv}$$

tendent vers zéro avec T. D'ailleurs, s'il en est ainsi, et puisque ρ tend vers l'infini les quantités

$$\frac{uB}{T},\quad \frac{Bv}{T},\quad \frac{uN}{T},\quad \frac{Nv}{T}$$

doivent aussi tendre vers l'infini et par suite aussi d'après (59), les quantités k, $u-k$, $B-k$, $N-u+k$ et donc aussi u, B, N, v. Reste à évaluer la quantité

$$A = \frac{B^B N^N u^u v^v}{T^T}\,\frac{1}{k^k (B-k)^{B-k} (u-k)^{u-k} (N-u+k)^{N-u+k}},$$

ou son logarithme

$$\begin{aligned} \mathcal{L}A = {} & B\,\mathcal{L}B + N\,\mathcal{L}N + u\,\mathcal{L}u + v\,\mathcal{L}v - T\,\mathcal{L}T \\ & - k\,\mathcal{L}\frac{uB}{T} - k\,\mathcal{L}\left(1+\frac{x\rho T}{uB}\right) \\ & - (B-k)\,\mathcal{L}\frac{Bv}{T} - (B-k)\,\mathcal{L}\left(1-\frac{x\rho T}{vB}\right) \\ & - (u-k)\,\mathcal{L}\frac{Nu}{T} - (u-k)\,\mathcal{L}\left(1-\frac{x\rho T}{Nu}\right) \\ & - (N-u+k)\,\mathcal{L}\frac{Nv}{T} - (N-u+k)\,\mathcal{L}\left(1+\frac{x\rho T}{Nv}\right), \end{aligned}$$

---

soit $\sqrt{\frac{BNuv}{T^2(T-1)}}$, mais comme il ne s'agira que de déterminer une expression asymptotique de P_k, il peut être préférable de remplacer cette expression par une expression équivalente plus simple, à choisir convenablement. C'est ce que la suite confirmera.

qui se réduit à

$$\mathcal{L}A = -k\,\mathcal{L}\left(1+\frac{x\rho T}{uB}\right) - (B-k)\,\mathcal{L}\left(1-\frac{x\rho T}{vB}\right)$$
$$- (u-k)\,\mathcal{L}\left(1-\frac{x\rho T}{Nu}\right) - (N-u+k)\,\mathcal{L}\left(1+\frac{x\rho T}{Nv}\right)$$

Utilisons comme dans le cas de Bernoulli l'égalité

$$\mathcal{L}(1+t) = t - t^2\left[\frac{1}{2} + \varepsilon(t)\right],$$

où $\varepsilon(t)$ tend vers zéro avec t. On aura

$$\mathcal{L}A = \frac{x\rho T^2}{uv}\left[\frac{u-k}{N} - \frac{k}{B}\right]$$
$$+ \frac{1}{2}(x\rho T)^2\left\{\frac{k}{u^2B^2} + \frac{B-k}{v^2B^2} + \frac{(u-k)}{N^2u^2} + \frac{(N-u+k)}{N^2v^2}\right\}$$
$$+ (x\rho T)^2\left\{\frac{k}{u^2B^2}\varepsilon_1 + \frac{B-k}{v^2B^2}\varepsilon_2 + \frac{u-k}{N^2u^2}\varepsilon_3 + \frac{N-u+k}{N^2v^2}\varepsilon_4\right\}$$
$$= U + \frac{V}{2} + \varepsilon V,$$

où $|\varepsilon|$ est au plus égal à un plus grand des nombres $|\varepsilon_1|, \ldots, |\varepsilon_4|$ et par suite tend vers zéro quand $\frac{x\rho T}{uB}, \ldots, \frac{x\rho T}{vN}$ tendent vers zéro, ce qui a certainement lieu avec $\frac{1}{T}$ d'après nos hypothèses.

Or, en remplaçant k, $u-k$, ... dans U et V au moyen des formules (59) on trouve

$$U = -\frac{x^2\rho^2T^3}{BNuv},$$

et $V = -U[1+\omega]$, avec

$$\omega = x\left[\frac{\rho T}{uB} + \frac{\rho T}{vN} - \frac{\rho T}{vB} - \frac{\rho T}{uN}\right].$$

Comme x est borné et comme chacun des termes du crochet tend, par hypothèse, vers zéro, il en est de même de ω. Or,

$$\mathcal{L}A = U\left[\frac{1}{2} - \left(\frac{\omega}{2} + \varepsilon + \varepsilon\omega\right)\right],$$

d'où

$$\mathcal{L}A \sim \frac{U}{2},$$

et, par suite,

$$\mathcal{L}A \sim -\frac{x^2\rho^2T^3}{2BNuv}$$

et

$$P_\lambda \sim \frac{1}{\sqrt{2\pi}}\sqrt{\frac{T^3}{BNuv}}\, e^{-\frac{\lambda^2}{2}\frac{\rho^2 T^3}{BNuv}}.$$

Ceci invite à prendre

$$\rho = \sqrt{\frac{2BNuv}{T^3}},$$

d'où

$$(60) \qquad P_\lambda \sim \frac{1}{\rho\sqrt{\pi}}\, e^{-\lambda^2},$$

de sorte qu'on a bien asymptotiquement, quand $T \to \infty$, une seconde loi de Laplace où la valeur moyenne de λ est $\frac{Bu}{T}$ et où l'écart quadratique moyen de λ est $\sqrt{\frac{BNuv}{T^3}}$. $\Big($Comme $\frac{BN}{T^2} \leqq \frac{1}{4}$ et $\frac{uv}{T^2} \leqq \frac{1}{4}$, cet écart est au plus égal à $\frac{\sqrt{T}}{4}$. Il est égal à $\frac{\sqrt{T}}{4}$ dans le cas considéré plus haut, $u = v = B = N$, comme nous l'avons déjà vu$\Big)$ D'ailleurs, quand $\rho = \sqrt{\frac{2BNuv}{T^3}}$, on a

$$\frac{\rho T}{uB} = 2\frac{Nv}{T^2}\frac{1}{\rho} < \frac{2}{\rho},$$

donc la condition que $\frac{\rho T}{uB}$ tende vers zéro se trouve réalisée si $\rho \to \infty$ et de même pour $\frac{\rho T}{Bv}$, $\frac{\rho T}{uN}$, $\frac{\rho T}{Nv}$. De même $\frac{\rho}{u}$ tend vers zéro puisque $\frac{\rho}{u} \leqq \frac{\rho T}{uB}$.

En résumé, la formule asymptotique (60) se trouve vérifiée quand T croît indéfiniment sous les seules conditions suivantes 1° que B, N, u, v, qui varieront en général avec T, varient de sorte que la quantité $\rho^2 = \frac{2BNuv}{T^3}$ tende vers l'infini avec T (1) $\Big($et comme B, N, u, v sont tous $\geqq \frac{\rho^2}{2}$, cela exige que B, N, u et v croîssent tous aussi au delà de toute limite$\Big)$, 2° qu'on applique

(1) La condition 1° sera remplie si le plus petit des nombres $\frac{B}{T}$ et $\frac{N}{T}$, et le plus petit des nombres $\frac{u}{T}$ et $\frac{v}{T}$, restent, quand T croît, supérieurs à un nombre positif fixe Mais cette dernière condition est plus stricte que (1°), comme on le voit en prenant (ce qui est possible pour T assez grand) :

$$N > 1 + T^{\frac{2}{3}} \geqq B > T^{\frac{2}{3}}, \qquad u > 1 + T^{\frac{2}{3}} \geqq v > T^{\frac{2}{3}}, \qquad \text{d'où} \qquad \frac{\rho^2}{2} = \frac{N}{T}\frac{u}{T}\frac{B}{T}v > \frac{T^{\frac{4}{3}}}{4}.$$

seulement la formule (57) aux valeurs de k, telles que la quantité

$$x = \left(k - \frac{nB}{T}\right)\frac{1}{\rho}$$

reste inférieure en valeur absolue à un nombre fixe quand T varie. [On pourrait d'ailleurs élargir cette condition (2°).]

Exemple de Lord Rayleigh modifié. — Dans l'exemple de Lord Rayleigh mentionné page 19, le nombre des états possibles est infini. Pour rester dans le cadre de ce Chapitre, modifions ce problème en ne considérant qu'un nombre fini d'états, d'autant que le problème ainsi modifié se présente dans quelques applications intéressantes.

Nous avons donc un point mobile A ne pouvant prendre qu'un nombre fini r de positions $A_1, A_2, \ldots, A_r$, et qui ne peut passer en une épreuve qu'à l'une des positions contiguës.

On peut imaginer ces positions, rangées dans l'ordre $A_1, \ldots, A_r$, soit sur un segment de droite $A_1 A_r$, soit sur un cercle. Dans les deux cas, on aura

$$\sum_k p_{jk} = 1$$

et

$$p_{jk} = 0 \qquad \text{pour} \quad j - k \neq \pm 1 \text{ quand } j \neq 1 \text{ et } j \neq r,$$

mais dans le cas rectiligne, on aura

$$p_{1k} = 0 \qquad \text{pour} \quad k \neq 2,$$
$$p_{rk} = 0 \qquad \text{pour} \quad k \neq r - 1,$$

et, par suite,

$$p_{12} = 1, \qquad p_{r,r-1} = 1,$$

tandis que dans le cas circulaire, on aura

$$p_{1k} = 0 \qquad \text{pour} \quad k \neq 2 \quad \text{et } k \neq r$$
$$p_{rk} = 0 \qquad \text{pour} \quad k \neq r - 1 \text{ et } k \neq 1.$$

Nous verrons (p. 126) que ces mêmes hypothèses sur les p_{jk} peuvent être interprétées d'une façon toute différente, dans un problème d'urnes.

Observons que $P_{jk}^{(n)}$ est la somme de produits de la forme $p_{ji_1}p_{i_1 i_2}\ldots p_{i_{n-1}k}$. Pour que $P_{jk}^{(n)}$ soit $\neq 0$, il faut que l'un au moins de ces produits soit $\neq 0$; pour un tel produit les indices $j, i_1, i_2, \ldots, i_{n-1}, k$ devront être de parités consécutives différentes dans le cas du mouvement rectiligne ou bien encore, si r est pair, dans le cas du mouvement circulaire. Ceci exige que $|k - j|$ et n soient de mêmes parités. Par conséquent $P_{jk}^{(n)} = 0$ si n et $|j - k|$ sont de parités distinctes. Alors quand j et k sont de même parité, $P_{jk}^{(n)}$ a pour limite zéro quand n croît par valeurs impaires et quand j et k sont de parités distinctes, $P_{jk}^{(n)}$ a pour limite zéro quand n croît par valeurs paires. Il en résulte

qu'on ne peut être dans le cas régulier. Car alors la limite unique P_k de $P^{(n)}_{jk}$ serait nulle quel que soit k et l'on ne pourrait avoir $\sum_k P_k = 1$. Nous traiterons ces deux cas singuliers plus loin, page 122

Exemple du cas positivement régulier — Le raisonnement ne s'applique pas lorsqu'il s'agit du mouvement circulaire avec un nombre impair de positions possibles. C'est ce cas que nous allons étudier maintenant, nous allons prouver que si les probabilités de passer en une épreuve d'une position à une position voisine sont toutes $\neq 0$, on se trouve dans le cas régulier et même positivement régulier.

On peut dans le mouvement circulaire résumer les hypothèses sur les p_{jk} sous la forme

$$p_{jk} \neq 0 \qquad \text{pour} \quad j - k = \pm 1 \qquad (\text{mod}\, r),$$

entendant par là que $j-k-1$ et $j-k+1$ sont des multiples (>0, <0 ou $=0$) de r et $p_{jk} = 0$ dans les autres cas.

Il est alors facile de voir qu'il existe un nombre entier ν et un rang k tel que les $P^{(\nu)}_{jk}$ soient tous $\neq 0$. Si, par exemple, on prend $\nu = r$ et $k = 1$, on voit qu'on peut aller de A_j en A_1 en r épreuves. Quand j est pair, on ira de A_j en A_{j-1}, ..., de A_2 en A_1, ce qui fait $j-1$ épreuves, puis on ira le nombre pair de fois, $r+1-j$, de A_1 en A_2, de A_2 en A_1, de A_1 en A_2, ..., de A_2 en A_1. Quand j est impair on ira de A_j à A_{j+1}, ..., de A_{r-1} à A_r, ce qui fait $r-j$ épreuves, puis on ira le nombre impair de fois, j, de A_r à A_1, de A_1 à A_r, de A_r à A_1, ..., de A_r à A_1. Par hypothèse tous ces mouvements ont des probabilités positives, donc

$$P^{(r)}_{j1} \neq 0 \qquad \text{pour} \quad j = 1, 2, \ldots, r$$

Nous sommes bien dans le cas régulier. Alors $P^{(n)}_{jk}$ tend vers une limite P_k indépendante de j. Les P_k sont donnés par les relations

$$P_k = \sum_i P_i \, p_{ik}.$$

Si $k \neq 1$ et $k \neq r$,

$$P_k = P_{k-1}\, p_{k-1,k} + P_{k+1}\, p_{k+1,k} \qquad (k = 2, \ldots, r-1) \tag{61}$$

et l'on a

$$P_1 = P_r \, p_{r1} + P_2 \, p_{21}, \tag{62}$$

$$P_r = P_{r-1}\, p_{r-1,r} + P_1\, p_{1r}. \tag{63}$$

Nous savons d'ailleurs qu'il y a une solution et une seule du système formé par ces équations et par l'équation $\sum_k P_k = 1$. On peut simplifier les notations

en posant

$$p_k = p_{k,k+1} \quad \text{et} \quad q_k = p_{k,k-1}, \quad q_1 = p_{1r}, \quad p_r = p_{r1}.$$

On a alors

$$(64) \qquad P_k = P_{k-1}\,p_{k-1} + P_{k+1}\,q_{k+1} \quad \text{pour} \quad k = 2, \ldots, r-1$$

et

$$(65) \qquad P_1 = P_r\,p_r + P_2\,q_2,$$

$$(66) \qquad P_r = P_{r-1}\,p_{r-1} + P_1\,q_1.$$

Par exemple, pour $r = 3$, on trouve

$$(67) \qquad \frac{P_1}{p_3 + q_3 q_2} = \frac{P_2}{p_1 + q_1 q_3} = \frac{P_3}{p_2 + q_2 q_1} = \frac{1}{2 + p_1 p_2 p_3 + q_1 q_2 q_3}.$$

Nous nous sommes placés dans le cas où $p_1, q_1, \ldots, p_r, q_r$ sont $\neq 0$. Les formules (64), (65), (66) montrent alors que $P_1, P_2, \ldots, P_r$ sont tous $\neq 0$. Car, si, par exemple, $P_k = 0$, on tirerait de (64) $P_{k-1} = 0$ et $P_{k+1} = 0$ et de proche en proche $P_1, P_2, \ldots$, seraient nuls, leur somme restant égale à 1.

Nous sommes bien dans le cas positivement régulier.

On observe que tous les P_k sont égaux — et par suite égaux à $\frac{1}{r}$ — quand les p_k, q_k sont tous égaux à $\frac{1}{2}$. Mais cette condition n'est pas nécessaire, car si les P_k sont égaux, on a

$$1 = p_{k-1} + q_{k+1} \quad \text{d'où} \quad p_{k-1} = p_{k+1} \text{ avec } p_r = p_2, \; p_{r-1} = p_1,$$

c'est-à-dire, r étant impair $(r = 2m+1)$

$$p_1 = p_3 = \ldots = p_r = p_2 = p_4 = \ldots = p_{2m},$$

donc tous les p_k sont égaux et de même tous les q_k. Ainsi pour qu'on soit dans le cas le plus régulier, celui où les $P_{jk}^{(n)}$ tendent vers une limite $\left(\text{égale à } \frac{1}{r}\right)$ indépendante de la position initiale et de la position finale, il faut et il suffit que la probabilité pour passer en une épreuve d'une position à la position voisine ait une valeur, p, indépendante de ces positions pour un sens de parcours déterminé. (La valeur correspondant à l'autre sens de parcours sera aussi indépendante de ces positions, mais sa valeur $q = 1 - p$ pourra être différente de la première, p, et par suite de $\frac{1}{2}$.)

Déplacement moyen. — Considérons le cas où les points $A_1, A_2, \ldots, A_r$

sont équidistants sur un même cercle et soit

$$l = \widehat{A_1A_2} = \widehat{A_2A_3} = \dots = \widehat{A_{r-1}A_r} = \widehat{A_rA_1}.$$

Le déplacement circulaire algébrique moyen effectué en n épreuves à partir de la position A_j est

$$\sum_k \widehat{A_jA_k}\, P_{jk}^{(n)}$$

(alors que la longueur totale parcourue a la valeur certaine nl.) Si l'on convenait de prendre pour $\widehat{A_jA_k}$ la détermination comprise de 0 à $(r-1)l$, le déplacement moyen serait

$$\sum_{k=j}^{k=r}(k-j)\,l\,P_{jk}^{(n)} + \sum_{k=1}^{k=j-1}(k-j+r)\,l\,P_{jk}^{(n)}$$

(la seconde somme disparaissant pour $j=1$.)

Si l'on prend la détermination de $-(r-1)\frac{l}{2}$ à $(r-1)\frac{l}{2}$, ce déplacement moyen sera

$$\sum_{k=j}^{k=r}\left(k-j-\frac{r-1}{2}\right)l\,P_{jk}^{(n)}+\sum_{k=1}^{j-1}\left(k-j+\frac{r+1}{2}\right)l\,P_{jk}^{(n)}$$

qui tend quand n croît vers

$$(68)\qquad l\left\{\sum_{k=1}^{r}(k-j)P_k+\frac{r+1}{2}\left[\sum_{k=1}^{k=j-1}P_k\right]-\frac{r-1}{2}\left[\sum_{k=j}^{k=r}P_k\right]\right\}.$$

Ce déplacement limite moyen, également compris de $-(r-1)\frac{l}{2}$ à $(r-1)\frac{l}{2}$, dépendra en général de la position initiale A_j. Toutefois, dans le cas particulier envisagé plus haut où les probabilités des déplacements de même sens sont égaux, cette valeur moyenne limite est égale à zéro.

La réciproque est vraie, on le voit en retranchant les valeurs de (68) pour $j=h$ et pour $j=h+1$. Si cette valeur moyenne limite est indépendante de l'état initial, on obtient $1-rP_h=0$, c'est-à-dire que P_h est indépendant de h.

Dispersion. — L'écart quadratique moyen correspondant à la même détermination du déplacement est la longueur μ_n telle que

$$\mu_n^2=\sum_{k=j}^{k=r}\left(k-j-\frac{r-1}{2}\right)^2 l^2 P_{jk}^{(n)}+\sum_{k=1}^{k=j-1}\left(k-j+\frac{r+1}{2}\right)^2 l^2 P_{jk}^{(n)}.$$

Quand n croît, il tend vers μ tel que

$$\mu^2 = l^2 \left[\sum_{k=j}^{k=r} \left(k - j - \frac{r-1}{2} \right)^2 P_k + \sum_{k=1}^{k=j-1} \left(k - j + \frac{r+1}{2} \right)^2 P_k \right]$$

Comme la moyenne du déplacement, il dépendra en général de j. Dans le cas le plus régulier, où l'on a $P_k = \frac{1}{r}$, on pourra écrire en posant $r = 2m+1$

$$\mu^2 = \frac{l^2}{r} \left\{ \sum_{u=-m}^{u=m+1-j} u^2 + \sum_{u=m+2-j}^{u=m} u^2 \right\}$$

$$= \frac{l^2}{r} \left\{ \sum_{-m}^{0} u^2 + \sum_{0}^{m} u^2 \right\} = \frac{2l^2}{r} \frac{m(m+1)(2m+1)}{6} = \frac{l^2}{12r}(r-1)(r+1)r,$$

$$\mu = \frac{l}{2}\sqrt{\frac{r^2-1}{3}} = \frac{1}{2}\sqrt{\frac{L^2-l^2}{3}}, \tag{69}$$

où L est la longueur de la circonférence.

On voit que pour de grandes valeurs de r, μ est équivalent à $\frac{rl}{2\sqrt{3}} = \frac{L}{2\sqrt{3}}$.

DEUXIÈME MÉTHODE

ÉTUDE DU CAS SINGULIER ET COMPLÉMENTS A L'ÉTUDE DU CAS RÉGULIER AU MOYEN DE L'EXPRESSION DES $P_{jk}^{(n)}$ EN FONCTION DE n.

Nécessité et avantages d'une nouvelle méthode. — Les résultats qui viennent d'être obtenus dans ce qui précède sont essentiellement ceux de Markoff, complétés seulement sur des points secondaires par ses successeurs. Ils sont, en outre, parmi les résultats les plus importants au point de vue pratique. Enfin la méthode de Markoff a l'avantage d'être à la fois simple et élémentaire.

Toutefois ces résultats sont incomplets et les compléments que nous allons indiquer seront obtenus par une extension convenable d'une autre méthode, celle de Poincaré-Romanovsky.

D'une part, en effet, la méthode de Markoff cesse d'être simple

quand on veut l'utiliser pour l'étude du comportement asymptotique des $P_{hk}^{(n)}$ dans le cas non régulier (dit encore singulier). D'autre part, s'il est vrai qu'elle conduit à des conditions suffisantes (pour qu'on soit dans le cas régulier) qui sont très simples (tous les $p_{hk} \neq 0$ ou bien : condition de M. Hostinský), par contre, la question de vérifier si la condition nécessaire et suffisante énoncée page 31 est satisfaite, est à peu près aussi difficile à traiter que le problème qu'elle prétend résoudre. Car elle consiste à s'assurer s'il existe une ligne du tableau des $P_{hk}^{(\nu)}$ qui soit $\neq 0$ pour une valeur convenable de ν. Comme cette valeur de ν n'est pas connue d'avance, il faudrait, par conséquent, effectuer une série — qui, à première vue, pourrait être illimitée — de calculs des $P_{hk}^{(\nu)}$, $(\nu = 1, 2, \ldots)$.

Ces divers inconvénients se trouvent évités par une extension convenable de la méthode de Poincaré. En même temps que celle-ci fournit des résultats plus complets que celle de Markoff, elle conduit aussi à des résultats plus généraux

M. Hadamard a déjà observé que la méthode de Poincaré, rendue un peu rébarbative par l'emploi de nombres complexes à n dimensions, devient plus naturelle dans le cas d'une suite continue d'événements en chaîne où ces nombres complexes sont remplacés par des « densités » de probabilités dont la signification est plus intuitive. Mais il faut reconnaître que la méthode de Markoff s'étend aussi à ce cas. Par contre, il y a une autre sorte d'extension qui est propre à celle de Poincaré.

La méthode et les résultats de Markoff utilisent en effet essentiellement à côté de la condition (I), les conditions (P) et (T) des pages 24, 25, sur lesquelles est fondé son principe de moyenne.

Or, la méthode de Poincaré fait surtout intervenir la condition d'itération (I). Celle-ci peut être interprétée comme un système d'équations aux différences finies. La discussion de ce système fait intervenir une équation dite équation en s ou équation séculaire. La méthode de Poincaré repose sur la considération de cette équation. Elle établit un lien avec les études purement algébriques de Frobenius. Plus tard, M. von Mises, les élèves de M. Hostinský et nous-même avons eu également recours à ces études.

Mais il est clair qu'on apportera plus de précision dans l'étude du comportement asymptotique des $P_{hk}^{(n)}$ si l'on en détermine l'expression directe en fonction de n. Or, on connaît l'expression générale des

solutions du système (I) d'équations aux différences considéré par Poincaré. C'est cette remarque qui a permis à M. Romanovsky [1] de compléter la méthode de Poincaré en utilisant cette expression dans le cas où les racines de l'équation en s sont simples. Les élèves de M. Hostinský et nous-même, avons étendu cette méthode au cas général où les racines sont simples ou multiples (Fréchet [147]). On obtient ainsi une discussion complète du problème de Markoff aussi bien dans le cas singulier que dans le cas régulier ([1]).

Extension hors du Calcul des Probabilités. — Tous ces résultats ont été obtenus dans l'hypothèse où la relation d'itération s'appliquait à des probabilités. Mais nous avons plus tard observé que les méthodes employées faisaient peu usage de cette hypothèse. De sorte qu'il nous a été possible (Fréchet [152]) de les étendre à des cas plus généraux. C'est donc encore un avantage de la méthode de Poincaré-Romanovsky de s'étendre, au delà du Calcul des Probabilités, à des problemes qui peuvent avoir leur intérêt propre, par exemple dans la théorie des équations intégrales (Fréchet [153]).

On peut donc commencer par étudier l'équation d'itération (I) dans le cas général où les quantités qui y figurent ne sont pas assujetties aux conditions (P), (T), spéciales au Calcul des Probabilités. On applique ensuite les résultats obtenus à ce cas spécial. Cette étude générale préalable constitue alors un problème d'algèbre et d'analyse dont il nous sera permis, dans ce Traité de Probabilité, de ne rappeler que les résultats et de reporter ceux-ci à la fin de ce volume dans les Notes A, B, C, page 256. Un certain nombre de ces résultats sont dus à Frobenius et à ses éleves, d'autres ont été obtenus par l'auteur de ce fascicule. On en trouvera les démonstrations complètes dans notre Mémoire de Brno (Fréchet [152]) où nous avons pu simplifier certaines démonstrations antérieures, de façon, par exemple, a éviter d'avoir recours à un théorème assez particulier de Minkowski sur les déterminants.

([1]) Nous avons complété plus loin les résultats de notre Mémoire de Pise, en empruntant à M. Dœblin [2] l'idée de faire intervenir la partie principale de $P_{ik}^{(n)}$ même dans le cas singulier, idée que nous appliquons dans la méthode algébrique actuelle après qu'il l'a utilisée dans la méthode directe que nous exposerons plus loin.

Les résultats propres au Calcul des Probabilités qui vont suivre, sont, la plupart, des applications immédiates de ces résultats algébriques généraux. Quelques-uns cependant nécessitent la démonstration préalable que nous allons donner d'une propriété des racines de module un de l'équation en s, définie plus loin, établie dans le cas spécial de l'application aux probabilités et qui n'est pas toujours vraie en dehors de ce cas.

Des remarques analogues à celles qui précèdent se présentent quand on passe au cas d'une suite continue d'épreuves.

Propriétés des racines de l'équation « en s ». — Les propriétés algébriques, indépendantes de la théorie des probabilités, qui ont été résumées dans les Notes A, B, C, mettent clairement en évidence l'importance, pour l'équation d'itération (I) de la page 24, de son *équation « en s »*, $\Delta(s) = 0$ avec

$$\Delta(s) \equiv \begin{vmatrix} p_{11}-s & p_{21} & & p_{r1} \\ p_{12} & p_{22}-s & & p_{r2} \\ & \cdot & & \\ p_{1r} & p_{2r} & & p_{rr}-s \end{vmatrix}$$

D'ailleurs, en ajoutant les lignes de $\Delta(s)$, on voit qu'en vertu de la condition (T_1) de la page 24, *cette équation « en s » a ici, nécessairement l'unité pour racine.*

Il résulte des théorèmes de Frobenius que les racines de cette équation sont toutes en module $\leqq 1$. Nous allons donner ici une démonstration de cette propriété qui va nous permettre de la préciser et de nous conduire à une propriété spéciale à celles des racines de $\Delta(s)$ qui sont de module 1. (L'attention portée à ces racines particulières s'explique par le rôle qu'elles jouent, dans le comportement asymptotique des $a_{ik}^{(n)}$ de la Note A, c'est-à-dire ici des $P_{ik}^{(n)}$).

I. Si $\Delta(s) = 0$, il y a un système, au moins, de solutions non toutes nulles $W_1, \ldots, W_r$ en $w_1, \ldots, w_r$ des équations

$$(1) \qquad s\,w_k = \sum_j p_{jk} w_j.$$

Or, que s, $W_1, \ldots, W_r$ soient réels ou complexes, on a, en dési-

gnant par R le plus grand des modules des W :

$$|s||W_k| \leqq \sum_j p_{kj} |W_j| \leqq R.$$

Pour l'une des valeurs de k, soit la valeur h, on a $|W_h| = R$, donc $|s| R \leqq R$, et comme $R \neq 0$,

$$(2) \qquad |s| \leqq 1.$$

C'est le premier théorème de Frobenius.

II. On peut le préciser de la façon suivante. On peut écrire (1) :

$$(s - p_{kk}) w_k = \sum_{j \neq k} p_{jk} w_j,$$

d'où

$$|s - p_{kk}||W_k| \leqq R \sum_{j \neq k} p_{jk} = R(1 - p_{kk})$$

En particulier

$$(3) \qquad |s - p_{hh}| \leqq 1 - p_{hh}.$$

L'inégalité (2) de Frobenius exprime que les affixes des racines de

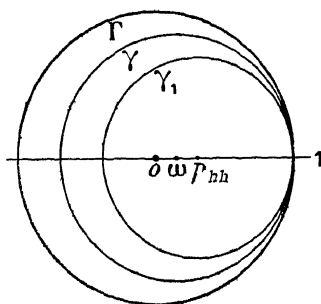



Fig. 1.

l'équation en s appartiennent, dans le plan complexe, à l'aire limitée par la circonférence Γ (*fig.* 1) de centre o passant par le point 1.

L'inégalité (3) exprime que ces racines appartiennent même à l'aire limitée par la circonférence γ_1 de centre p_{hh} passant par 1. Seulement, pour l'utiliser directement, il faudrait savoir pour quelle valeur de h, on a $|W_h| = R$. Mais appelons ω le plus petit des

nombres p_{11}, .., p_{kk}, ..., p_{ll}. La circonférence γ de centre ω passant par 1 est comprise dans Γ, mais comprend γ_1, de sorte que l'inégalité (3) a pour conséquence

$$(4) \qquad |s-\omega| \leqq 1-\omega$$

Nous avions d'abord démontré (Fréchet [152], p. 23) l'inégalité nouvelle (4) en utilisant une démonstration algébrique d'un intéressant théorème de M. Tambs-Lyche sur les déterminants (démontré d'abord par lui par une voie non algébrique). Nous venons maintenant de donner une seconde démonstration, directe et très simple, de cette inégalité (4). Celle-ci se réduit à l'inégalité (2) quand $\omega = 0$; elle est plus précise que l'inégalité de Frobenius, qui en devient une conséquence, quand $\omega \neq 0$. L'inégalité (4) permet de conclure que : quand les éléments p_{kk} de la diagonale principale de D sont tous $\neq 0$, il n'y a pas de racine de module 1 de l'équation en s autre que l'unité. Ce résultat sera étendu plus loin (p. 108).

III. *Racines de module* 1. — La solution $s = 1$ a pour module 1. s'il y en a d'autres, que peut-on dire ? Si $|s| = 1$ et $|W_h| = R$, on aura

$$R = \left| \sum_j p_{hj} W_j \right|$$

Le second membre est la distance à l'origine, du centre c des moyennes distances de points tous situés à l'intérieur ou sur le contour du cercle $|z| = R$. Il faudrait que ce centre fût aussi sur le contour de ce cercle. Si l'on prend les moments par rapport à la tangente à la circonférence au point c, des poids p_{h1}, p_{h2}, ..., p_{hl}, appliqués en W_1, W_2, ..., la somme de ces moments, tous $\geqq 0$, devra être nulle. Chacun d'eux devant alors être nul, il faudra que les poids p_{hj} des points W_j distincts de c soient nuls ([1]). Donc

$$\sum_j p_{hj} W_j = c \sum_j p_{hj} = c,$$

([1]) On peut aussi raisonner algébriquement

En posant $c = \sum_j p_{hj} W_j$, on a $|c| = R$, donc $c \neq 0$ et par suite $\sum_j p_{hj}\left(1 - \frac{W_j}{c}\right) = 0$

En posant $\frac{W_j}{c} = \alpha_j + i\beta_j$, on aura $|\alpha_j| \leqq \left|\frac{W_j}{c}\right| \leqq 1$ et $\sum_j p_{hj}(1-\alpha_j) = 0$. Dans cette

et par suite $s W_h = c$. Comme la somme $\sum_j p_{hj} = 1$, l'un au moins des p_{hj}, $p_{hh'}$, par exemple, est $\neq 0$ et alors $W_{h'}$ est en c. D'ailleurs W_h étant sur la circonférence ou bien c est en W_h ou bien $c \neq W_h$ et $p_{hh} = 0$. Or, dans le premier cas, on aurait $s W_h = W_h$ avec $W_h \neq 0$ d'où $s = 1$, cas que nous écartons. Ainsi $p_{hh} = 0$. On retrouve ce résultat établi autrement plus haut, que le cas examiné ne peut se présenter si la diagonale principale de **D** est positive.

Nous voyons de plus que $W_{h'} = c \neq W_h$ avec $|W_{h'}| = |c| = R$. Il y a donc deux inconnues, de valeurs distinctes : W_h et $W_{h'}$ dont les modules sont égaux à R. Soient W_h, $W_{h'}$, $W_{h''}$, ... les solutions, de valeurs distinctes ou non, dont le module est égale à R, on voit comme plus haut que

$$p_{hh} = p_{h'h'} = p_{h''h''} = \ldots = 0$$

Ainsi, *il ne peut y avoir de racine de module* 1 *de l'équation « en s », qui soit autre que* 1, *que si la diagonale principale comporte au moins deux éléments nuls*. Ce résultat sera encore précisé, page 186.

Parmi ces solutions de module R, distinguons celles dont les valeurs sont différentes et désignons-les par W_{h_1}, W_{h_2}, ..., W_{h_m} (il y en a au moins deux : W_h et $W_{h'}$). Pour chacune, W_{h_1}, par exemple, on a une égalité de la forme $s W_{h_1} = W_{h'_1}$. Les $W_{h'_1}$, $W_{h'_2}$, ... ayant alors pour module R sont prises parmi les W_{h_1}, W_{h_2}, ..., W_{h_m}. De plus, elles sont nécessairement distinctes puisque $s \neq 0$. Les valeurs $W_{h'_1}$, $W_{h'_2}$, ..., $W_{h'_m}$ sont donc les mêmes que W_{h_1}, W_{h_2}, ..., W_{h_m}, mais dans un autre ordre. Dès lors, en multipliant ces équations, on a

$$s^m = 1 \qquad \text{avec} \quad m \leqq r$$

Ainsi le *module d'une racine de* $\Delta(s)$ *ne peut être égal à l'unité que si cette racine est racine d'une équation binome* de la forme $s^m = 1$ où m est un entier positif $\leqq r$. D'ailleurs, les racines de module 1 de $\Delta(s)$, étant racines d'équations binomes de degré m_1, m_2, ..., sont alors racines d'une même équation binome $s^N = 1$ où N est un multiple commun de m_1, m_2, ...

---

somme, tous les termes sont $\geqq 0$, donc, pour $p_{hj} \neq 0$, on a $1 = \alpha_j \leq \sqrt{\alpha_j^2 + \beta_j^2} \leq 1$, d'où $\beta_j = 0$ et par suite $W_j = c$.

Inversement, *de telles racines de $\Delta(s)$ peuvent effectivement se présenter*, comme dans le cas où le déterminant D des p_{jk} serait

$$\begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix}.$$

Dans ce cas, $\Delta(s) = 1 - s^3$ a pour racines les trois racines cubiques de l'unité.

a. — Comportement des probabilités itérées

Les probabilités $P_{jk}^{(n)}$ convergent toujours en moyenne arithmétique. Leurs divers modes de comportement asymptotique quand n croît. — Sans beaucoup plus de peine que si nous nous étions limités au problème de Markoff, nous avons pu faire dans la Note A une étude assez complète du comportement asymptotique des $a_{jk}^{(n)}$. En appliquant ces résultats au problème des probabilités en chaîne, on précise encore les résultats obtenus Il suffit de tenir compte dans les énoncés qu'on se trouve nécessairement dans « le cas borné » défini page 259, que la condition (T_1) est vérifiée d'elle-même et de faire intervenir les propriétés de l'équation « en s », spéciales au cas des probabilités, qui viennent d'être énoncées. On a alors les propositions suivantes :

Dans le problème des probabilités en chaîne : *les probabilités itérées $P_{jk}^{(n)}$ convergent toujours en moyenne arithmétique* (définition p. 71 et 259) *vers des limites Π_{jk}. Ces limites généralisées Π_{jk}, nécessairement $\geqq 0$, ne sont pas toutes nulles, et même on a*

$$(5) \qquad \sum_{k=1}^{k=r} \Pi_{jk} = 1 \qquad (j = 1, 2, \ldots, r).$$

La moyenne

$$\Pi_{jk}^{(n)} = \frac{1}{n}[P_{jk}^{(1)} + \ldots + P_{jk}^{(n)}]$$

ne diffère donc de sa limite Π_{jk} que par un infiniment petit avec $\frac{1}{n}$: *cet infiniment petit est au moins du premier ordre.*

Plus précisément, $n[\Pi_{jk}^{(n)} - \Pi_{jk}]$ est borné quand n varie ; et même,

dans le cas où les $P_{jk}^{(n)}$ *convergent au sens ordinaire, ce produit a une limite* s_{jk}, *qui est la somme de la série absolument convergente*

$$s_{jk} = \sum_{t=1}^{t=\infty} [P_{jk}^{(t)} - \Pi_{jk}]$$

Le raisonnement fait aux pages 46-47 dans le cas régulier s'étend a ce cas et montre qu'on a

$$(6)\quad \begin{cases} s_{ik} - p_{ik} + P_{ik} = \sum_j s_{ij}\, p_{jk} = \sum_j p_{ij}\, s_{jk} \\ \sum_j s_{ij} = 0, \qquad \sum_i P_{ji}\, s_{ih} = 0 \qquad \text{et aussi} \qquad \sum_i s_{ji}\, P_{ih} = 0 \end{cases}$$

Revenons au cas général.

Les valeurs de Π_{jk} vérifient en outre de (5) les relations

$$(7)\qquad \Pi_{jk} = \sum_i \Pi_{ji}\, p_{ik} = \sum_i p_{ji}\, \Pi_{ik}$$

et

$$(8)\qquad \Pi_{jk} = \sum_i \Pi_{ji}\, \Pi_{ik}$$

D'apres (5) et (7), on voit que : pour chaque valeur de j, Π_{j1}, Π_{j2}, ..., Π_{jk} *forment un système de solutions du système d'équations* ($\mathcal{E}$) de la page 44.

L'expression de Π_{jk} *est de la forme*

$$(9)\qquad \Pi_{jk} = \sum_{g=1}^{g=\rho} U_{jg} V_{kg},$$

où U_{1g}, ..., U_{rg} *est un système* S_g *de solutions non toutes nulles des équations*

$$(10)\qquad u_j = \sum_{i=1}^{i=r} p_{ji}\, u_i \qquad (j = 1, \ldots, r),$$

où V_{1g}, ..., V_{rg} *est un système* S'_g *de solutions non toutes nulles des équations*

$$(11)\qquad y_k = \sum_{i=1}^{i=r} y_i\, p_{ik} \qquad (k = 1, \ldots, r),$$

où $S_1, \ldots, S_\rho$ est un système arbitraire de solutions S_g linéairement indépendantes en nombre maximum $\rho\,(\leqq r)$, des équations (10), *et où $S'_1, \ldots, S'_\rho$ est le système parfaitement déterminé de solutions S'_g de* (11) *qui forme avec le système des S_g un système biorthogonal et normé par rapport aux seconds indices*, c'est-à-dire tel que

$$\sum_i U_{ig} V_{ih} = \delta_{gh} \qquad \text{avec} \quad \delta_{gh} = \begin{cases} 1 & \text{si } g = h, \\ 0 & \text{si } g \neq h \end{cases}$$

On peut ajouter qu'ici, la condition (T_1) étant réalisée, les équations (10) sont vérifiées pour $x_1 = \ldots = x_r = 1$, par suite, on peut prendre $U_{i1} = 1$ pour $i = 1, \ldots, r$.

Pour que les limites généralisées Π_{jk} soient indépendantes du premier indice j [1], *il faut et il suffit qu'il n'existe qu'un système de solutions du système d'équations*

$$(\mathcal{E}) \quad \begin{cases} (12) \quad y_k = \sum_{i=1}^{i=r} y_i p_{ik} \qquad (k = 1, \ldots, r), \\ (13) \quad \sum_{i=1}^{i=r} y_i = 1, \end{cases}$$

et dans ce cas, les $\Pi_k = \Pi_{jk}$ ont précisément pour valeurs ce système de solutions [2], *ou encore il faut et il suffit que l'unité soit racine simple de l'équation en s, $\Delta(s) = 0$.*

Pour que les $P_{jk}^{(n)}$ convergent au sens ordinaire vers des limites respectives, il faut et il suffit que $\Delta(s)$ n'ait pas de racine de module 1 *autre que* 1. Et, dans ce cas, les limites au sens ordinaire des $P_{jk}^{(n)}$ étant respectivement égales aux limites généralisées Π_{jk}, sont déterminées comme plus haut sous la forme (9).

---

(1) Dans nos publications précédentes nous avons appelé cas *semi-régulier* le cas où les $P_{jk}^{(n)}$ convergent au sens ordinaire. En raison de l'importance du principe ergodique, il convient mieux d'appeler ainsi le cas où la limite généralisée de $P_{jk}^{(n)}$ est indépendante de j.

(2) On observera en se reportant à l'exemple de la page 45 que c'est notre introduction des limites généralisées qui permet de donner une signification à la solution unique du système $(\mathcal{E})$, en dehors du cas régulier.

Pour qu'on soit dans le cas régulier, il faut et il suffit : 1° que l'équation en s n'ait pas de racine de module 1 autre que l'unité, et 2° que celle-ci soit racine simple de $\Delta(s)$.

La condition nécessaire a été publiée par M. Konečný [1], la condition suffisante par M. Kaucky [1]. En arrivant indépendamment à ces deux résultats et par d'autres méthodes (Fréchet [47]), nous avons pu, en outre, *grâce à l'introduction de la limite en moyenne*, donner leurs significations respectives aux conditions 1° et 2°.

Une autre forme des mêmes conditions est la suivante :

Pour qu'on soit dans le cas régulier, il faut et il suffit : 1° que l'équation en s n'ait pas de racine de module 1 autre que l'unité; 2° qu'il n'existe qu'un système de solutions du système ($\mathcal{E}$) d'équations (et, dans ce cas, les $P_k = P_{jk}$ ont précisément pour valeurs ce système de solutions).

En particulier, *pour qu'on soit dans le cas positivement régulier, il faut et il suffit : 1° que l'équation en s n'ait pas de racine de module 1 autre que l'unité, 2° qu'il n'existe qu'un système de solutions du système ($\mathcal{E}$) d'équations, 3° que ces solutions qui sont nécessairement $\geqq 0$, soient toutes $\neq 0$.*

Pour qu'on soit dans le cas le plus régulier, il faut et il suffit : 1° que la condition (T'_1), $\sum_i p_{ik} = 1$ soit réalisée, 2° que l'équation en s n'ait pas de racine de module 1 autre que l'unité, 3° que celle-ci soit racine simple de $\Delta(s)$. Et alors *les limites P_{jk} sont toutes égales à $\frac{1}{r}$.*

En particulier, dans le cas symétrique ($p_{ik} = p_{ki}$), on ne peut être dans le cas régulier que si l'on est dans le cas le plus régulier.

On notera que toutes ces conditions présentent un avantage sur celles déduites (p. 31 et 39) de la méthode de Markoff. C'est qu'on peut s'assurer de leur réalisation directement sur les p_{jk} et sur l'équation en s que celles-ci déterminent, alors que la méthode de Markoff, très commode si une ligne des p_{jk} est $\neq 0$, oblige, dans le cas contraire, à former un nombre, non déterminé d'avance, de probabilités itérées.

Toutes les propriétés des $P_{jk}^{(n)}$ qui viennent d'être énoncées reposent d'abord sur le résultat suivant (Note A, p. 258 et 272) :

Les $P_{jk}^{(n)}$ *ont, pour* n *assez grand* (en tout cas pour $n > r$) *une expression de la forme*

$$P_{jk}^{(n)} = \sum_g (\sigma_g)^n R_{jkg}(n), \tag{12}$$

où $\sigma_1, \sigma_2, \ldots$ *sont les racines distinctes de l'équation en* s, $\Delta(s) = 0$, *et où* $R_{jkg}(n)$ *est un polynome en* n (de degré inférieur à l'ordre de multiplicité de σ_g comme racine de $\Delta(s)$].

Ajoutons enfin, en combinant les résultats des pages 108 et 260, une propriété qui a été démontrée d'abord dans le cas particulier où le déterminant D des p_{ik} est « indécomposable » (*voir* p. 168) par M. von Mises, d'ailleurs par une méthode tout à fait différente, et qui se trouve maintenant établie dans le cas le plus général :

Les probabilités itérées $P_{jk}^{(n)}$ *sont toujours des fonctions asymptotiquement périodiques de* n —, c'est-à-dire sont chacune somme d'une fonction périodique $\varpi_{jk}^{(n)}$ de n et d'une quantité $\varepsilon_{jk}^{(n)}$ infiniment petite avec $\frac{1}{n}$. *On peut prendre comme période commune des* $\varpi_{jk}^{(n)}$, *le degré de l'équation binome* $s^N - 1 = 0$, *à laquelle satisfont nécessairement les racines de module* 1 *de* $\Delta(s)$.

On peut ajouter que $\varepsilon_{jk}(n)$ *converge « exponentiellement »* (p. 30) *vers zéro* avec $\frac{1}{n}$. On peut même préciser encore : $\varepsilon_{jk}^{(n)}$ est la partie de la somme $\sum_g$ ci-dessus correspondant aux $\sigma_g \neq 1$ et de module 1. De sorte que la précision définitive à apporter aux énoncés de la page 30 est la suivante : $\varepsilon_{jk}^{(n)}$ converge au moins aussi rapidement vers zéro qu'une progression géométrique dont la raison est l'un quelconque des nombres < 1 et supérieurs au plus grand module des σ_g de modules < 1.

$\varpi_{jk}^{(n)}$ étant l'autre partie de la somme $\sum_g$, on sait (p. 259) que tous les $R_{jkg}(n)$ correspondants se réduisent à des constantes, que pour $\sigma_g = 1$, on a $R_{jkg} = \Pi_{jk}$ et que pour ces σ_g, on a $(\sigma_g)^N = 1$, *de sorte* que

$$\varpi_{jk}^{(n)} = \Pi_{jk} + \sum_g{}' e^{\frac{2\pi i n m_g}{N}} R_{jkg}, \tag{13}$$

où les m_g et N sont des entiers.

Dans le cas où les $P_{jk}^{(n)}$ convergent au sens ordinaire, la période de $\varpi_{jk}^{(n)}$ est 1. On peut appeler ce cas, le cas *non-oscillant*. Dans le cas symétrique, la période est 2 ou 1 suivant que -1 est, ou n'est pas, racine de $\Delta(s)$.

Dans le cas (symétrique ou non) où toutes les racines sont simples et $\neq 0$, on peut écrire explicitement l'expression de $P_{jk}^{(n)}$ sous la forme :

$$P_{jk}^{(n)} = \varphi_{j1}\psi_{k1} + \sum_{g>1} \frac{\varphi_{jg}\psi_{kg}}{(\lambda_g)^n}. \tag{14}$$

(Cette forme rappelle l'expression classique des noyaux symétriques itérées en prenant $\lambda_1 = 1$.) On montre dans la Note A, à la page 264, que la détermination des coefficients, — indépendants de n —, φ et ψ, s'obtient par la résolution successive de deux systèmes d'équations du premier degré et que les φ et ψ forment un système biorthonormé aussi bien par rapport au premier que par rapport au second indice, c'est-à-dire que

$$\sum_g \varphi_{jg}\psi_{kg} = \delta_{jk} \qquad \text{et} \qquad \sum_i \varphi_{ig}\psi_{ig'} = \delta_{gg'}.$$

D'autre part les λ_g sont les inverses des racines de l'équation en s.
On a d'ailleurs pour les φ

$$\varphi_{jg} = \lambda_g \sum_i p_{ji}\varphi_{ig},$$

de sorte que pour $g = 1$, et $\lambda_g = 1$, on peut prendre les $\varphi_{i1} = 1$, d'où la formule de M. Romanovsky

$$P_{jk}^{(n)} = \Pi_k + \sum_{g>1} \frac{\varphi_{jg}\psi_{kg}}{(\lambda_g)^n}, \tag{15}$$

car ψ_{k1} est évidemment la limite en moyenne Π_k de $P_{jk}^{(n)}$.

Calcul des probabilités limites. — Dans sa Thèse (en roumain), M. Mihoc [1] a indiqué une expression simple des probabilités limites P_j dans le cas régulier. Sa démonstration s'appliquant d'ailleurs identiquement dans le cas semi-régulier au calcul des limites en moyenne π_j, va être indiquée dans ce cas.

On part d'une identité simple obtenue en remplaçant dans $\Delta(s)$ la $k^{ième}$ ligre

par les sommes de toutes les lignes de $\Delta(s)$. On en tire

$$\frac{\Delta(s)-\Delta(1)}{s-1} = -\begin{vmatrix} p_{11}-s & p_{21} & \cdot & p_{r1} \\ \cdot & & & \cdot \\ 1 & 1 & & 1 \\ \cdot & & & \\ p_{1r} & p_{2r} & \cdot & p_{rr}-s \end{vmatrix};$$

d'où en désignant par Δ_{jk} le coefficient du terme situé dans la j^{ieme} colonne et la k^{ieme} ligne dans le développement de $\Delta(1)$

$$-(\Delta'_s)_{s=1} = \Delta_{1k}+\Delta_{2k}+\ldots+\Delta_{rk}.$$

Il en résulte d'abord que dans tous les cas le second membre a une valeur A indépendante du rang k de la ligne.

Dans le cas semi-régulier, 1 étant racine simple de $\Delta(s)$, on voit que A est $\neq 0$. Or, A est égal au déterminant des coefficients dans le système $(\mathcal{E}_k)$ obtenu en supprimant la k^{ieme} équation du système $(\mathcal{E})$ de la page 111.

Il en résulte une nouvelle preuve du fait que ce système $(\mathcal{E})$ a un seul système de solutions dans le cas où 1 est racine simple de $\Delta(s)$. Et dans ce cas il suffit, pour trouver les solutions $\pi_1, \pi_2, \ldots$ de ce système, d'appliquer la règle de Cramer au système $\mathcal{E}_k$. On a alors évidemment

$$\frac{\pi_1}{\Delta_{1k}} = \frac{\pi_2}{\Delta_{2k}} = \ldots = \frac{\pi_r}{\Delta_{rk}} = \frac{1}{\Delta_{1k}+\ldots+\Delta_{rk}},$$

d'où

$$\pi_1 = \frac{\Delta_{1k}}{A}, \quad \ldots, \quad \pi_r = \frac{\Delta_{rk}}{A}.$$

Par suite, dans le cas semi-régulier, $\Delta_{1k}, \ldots, \Delta_{rk}$ sont indépendants du second indice k. On peut alors poser $\Delta_{jk} = \Delta_j$ et l'on obtient ainsi *les formules de M. Mihoc*

$$(\alpha) \quad \pi_1 = \frac{\Delta_1}{\Delta_1+\ldots+\Delta_r}, \quad \ldots, \quad \pi_j = \frac{\Delta_j}{\Delta_1+\ldots+\Delta_r}, \quad \ldots, \quad \pi_r = \frac{\Delta_r}{\Delta_1+\ldots+\Delta_r}$$

valables dans le cas semi-régulier, qui est précisément celui où

$$\Delta_1+\ldots+\Delta_r \neq 0.$$

Remarque. — En développant $\Delta(1)$ suivant la k^{ieme} ligne, on a

$$0 = \Delta(1) = p_{1k}\Delta_1+\ldots+(p_{kk}-1)\Delta_k+\ldots+p_{rk}\Delta_r,$$

d'où

$$\Delta_k = p_{1k}\Delta_1+\ldots+p_{rk}\Delta_r$$

et

$$(\beta) \quad \pi_k = \frac{p_{1k}\Delta_1+\ldots+p_{rk}\Delta_r}{\Delta_1+\ldots+\Delta_r}.$$

Nous savions déjà d'après (7) que π_k est une moyenne pondérée des $p_{1k}, \ldots, p_{rk}$. Or, d'après (α), les Δ_j sont tous de même signe (tous $\geqq 0$ ou tous $\leqq 0$). M. Mihoc a conclu de la formule (β) qu'on peut prendre pour poids les quantités connues $\Delta_1, \ldots, \Delta_r$.

Calcul des s_{jk}. — Revenons en arrière; on a noté que $(\Pi_{jk}^{(n)} - \Pi_{jk})$ est infiniment petit au moins du premier ordre. Plus précisément, l'expression $R_{jk}^{(n)} = n(\Pi_{jk}^{(n)} - \Pi_{jk})$ est bornée. On peut préciser son comportement asymptotique. Il résulte de la Note B, page 277, que $R_{jk}^{(n)} = \sum_{t=1}^{t=n} [P_{jk}^{(t)} - \Pi_{jk}]$ est une fonction asymptotiquement périodique de n. Dès lors, elle converge en moyenne.

Dans le cas non-oscillant, $R_{kj}^{(n)}$ converge au sens ordinaire, et par suite aussi en moyenne, vers s_{jk}. On peut alors, avec M. Dœblin, généraliser la définition de s_{jk}. Nous continuerons à appeler dans le cas oscillant s_{jk} la limite en moyenne de $R_{jk}^{(n)}$, c'est-à-dire que nous pourrons écrire par définition, dans tous les cas,

$$s_{jk} = \sum_{t=1}^{t=\infty}{}' (P_{jk}^{(t)} - \Pi_{jk}),$$

Σ' indiquant qu'il s'agit de la somme généralisée, en moyenne. On prouve (Note B) qu'on a les relations, analogues aux formules (6)

$$s_{jk} - (p_{jk} - \Pi_{jk}) = \sum_i s_{ji} p_{ik} = \sum_i p_{ji} s_{ik},$$

$$\sum_i s_{ji} \Pi_{ik} = 0,$$

$$\sum_i \Pi_{ji} s_{ik} = 0,$$

d'où l'on déduit par addition de l'avant-dernière, écrite pour $k = 1, \ldots, r$,

$$\sum_i s_{ji} = 0.$$

Dès lors $s_{j1}, s_{j2}, \ldots, s_{jr}$ sont solutions du système :

$$(\mathcal{E}''_j) \qquad \begin{cases} z_k - \sum_i p_{ik} z_i = p_{jk} - \Pi_{jk} & (k = 1, \ldots, r), \\ \sum_i z_i = 0. \end{cases}$$

On peut réduire le système à r équations en supprimant l'une des r premières, conséquence des autres comme on le voit par addition On sait que dans le cas semi-régulier, le système ($\mathcal{E}$) de la page 111, a un seul système de solutions; or, après réduction à r équations il a le même déterminant des coefficients que le systeme actuel ($\mathcal{E}''_j$). Dès lors, *dans le cas semi-régulier* les $s_{j1}, s_{j2}, \ldots$ forment le système unique de solutions du système ($\mathcal{E}''$).

Discussion dans le cas où il n'y a que deux états possibles — Employons les notations de la page 90. Alors

$$\Delta(s) = \begin{vmatrix} 1-a-s & b \\ a & 1-b-s \end{vmatrix} = \begin{vmatrix} 1-a-s & b \\ 1-s & 1-s \end{vmatrix}$$

en ajoutant les lignes. D'où

$$\Delta(s) = (1-s)(1-a-b-s)$$

Les deux racines sont $s_1 = 1$, $s_2 = 1-a-b$, et comme $0 \leqq {a \atop b} \leqq 1$ on vérifie que $|s_2| \leqq 1$.

I. Pour qu'on soit dans le cas *semi-régulier*, il faut que l'unité ne ne soit pas racine multiple, c'est-à-dire qu'on n'ait pas $a = b = 0$. On sera dans le cas singulier non oscillant si $a = b = 0$, c'est-à-dire si $p_{11} = 1 = p_{22}$, $p_{12} = p_{21} = 0$. C'est le cas où une épreuve ne peut produire un changement d'état ou seulement avec une probabilité nulle. Il en est donc de même après n épreuves. Les $P_{ik}^{(n)}$ sont constants, mais dépendent de l'état initial E_i.

II. Pour que les $P_{jk}^{(n)}$ convergent au sens ordinaire, il faut qu'il n'y ait pas de racine de module 1 qui soit $\neq 1$. Le cas contraire : le cas oscillant, est donc celui où l'on a

$$1-a-b = \pm 1 \qquad \text{et} \qquad 1-a-b \neq 1,$$

c'est-à-dire que $1 - a - b$ doit être égal à -1 ; d'où

$$0 \leqq 1 - a = -(1 - b) \leqq 0$$

et, par suite,

$$(1 - a) = -(1 - b) = 0,$$

c'est-à-dire $a = 1$, $b = 1$,

$$p_{11} = 0, \qquad p_{22} = 0, \qquad p_{12} = p_{21} = 1.$$

Dans ce cas, il y a, à chaque épreuve, changement d'état sûrement (ou avec une probabilité égale à 1). Les $P_{ik}^{(n)}$ sont alternativement 0 ou 1.

Dans tous les autres cas, les $P_{ik}^{(n)}$ tendent vers des limites déterminées.

Pour qu'on soit dans le cas régulier, il faut en outre que la racine $s_1 = 1$ soit simple, c'est-à-dire que $s_2 \neq 1$ ou $1 - a - b \neq 1$ ou $a + b \neq 0$.

Pour qu'on soit dans le cas régulier, il faut donc que a et b ne soient pas tous deux nuls ou tous deux égaux à l'unité. Dans ce cas régulier, on a vu (p. 90) que les $P_{ik}^{(n)}$ tendent vers des limites P_k telles que

$$P_1 = \frac{b}{a + b}, \qquad P_2 = \frac{a}{a + b},$$

et pour qu'on soit dans le cas positivement régulier, il faut encore que a et b soient $\neq 0$.

Enfin, pour qu'on soit dans le cas le plus régulier, celui où $P_1 = P_2$, il faut que $a = b$ et leur valeur commune doit, comme plus haut, être $\neq 0$ et $\neq 1$. Alors on a $P_1 = P_2 = \frac{1}{2}$.

On a ainsi le tableau suivant :

Dans tous les cas $0 \leqq {a \atop b} \leqq 1$.

I. $0 < a = b < 1$, cas le plus régulier, $P_1 = P_2 = \frac{1}{2}$.

II. $0 < {a \atop \text{et} \atop b}$ et ${a \atop \text{ou} \atop b} < 1$, cas positivement régulier

$$P_1 = \frac{b}{a + b}, \qquad P_2 = \frac{a}{a + b}.$$

III. $a=0$ et $b\neq 0$ ou $b=0$ et $a\neq 0$, cas régulier sans être positivement régulier. Si par exemple $a=0\neq b$, $P_1=1$, $P_2=0$.

En outre, on a observé, page 92, que, si $a+b=1$, — ce qui peut avoir lieu dans les cas I, II, III —, on a $p_{1k}=p_{2k}$ pour $k=1, 2$ et, en appelant P_k leur valeur commune, on a $P_{ik}^{(n)}=P_k$ quel que soit n.

IV. $a=b=1$, cas oscillant : $P_{ik}^{(n)}=$ alternativement 1 ou 0 quand n croît.

Cependant, dans ce cas, $\Pi_{ik}^{(n)}$ tend vers $\frac{1}{2}$. De sorte qu'on est encore dans le cas semi-régulier comme dans les cas précédents.

V. $a=b=0$, cas singulier non oscillant

$$P_{ik}^{(n)}=P_{ik} \qquad \text{avec} \qquad P_{1k}\neq P_{2k}.$$

Ici, la limite P_{ik} de $\pi_{ik}^{(n)}$ dépend de i.

On peut facilement former l'expression générale des $P_{hk}^{(n)}$ en fonction de n. C'est fait plus haut dans le cas V : alors $P_{hk}^{(n)}$ est indépendant de n. Dans tous les autres cas, $1-a-b\neq 1$ et $s_2\neq s_1$. Les deux racines s_1, s_2 étant distinctes, l'expression cherchée sera, pour $n>2$, de la forme (Note A)

$$P_{hk}^{(n)}=\omega_{hk}^{(1)}+(s_2)^n\omega_{hk}^{(2)},$$

où $s_2=1-a-b$. Si $a+b=1$, $s_2=0$, les $P_{hk}^{(n)}$ sont des constantes quand n varie. Il reste donc le cas où $s_2\neq 0$, c'est-à-dire où $a+b<1$. On a montré, page 112, comment, pour r quelconque, on peut calculer $P_{hk}^{(n)}$ dans le cas où $\Delta(s)$ a ses racines distinctes et $\neq 0$. En appliquant ce résultat pour calculer explicitement les coefficients $\omega_{hk}^{(g)}$ dans le cas actuel où $r=2$ et $a+b\neq 1$, on trouve $P_{hk}^{(n)}$ en fonction de n. On a alors, d'après (15) de la page 114,

$$P_{hk}^{(n)}=P_k+\varphi_{h2}\psi_{k2}(1-a-b)^n,$$

qu'on peut écrire

$$P_{hk}^{(n)}=P_k+\beta_h\gamma_k(1-a-b)^n,$$

avec

$$P_1=p_{11}P_1+p_{21}P_2 \qquad \text{et} \qquad P_1+P_2=1,$$

d'où, comme page 90,

$$P_1=\frac{b}{a+b}, \qquad P_2=\frac{a}{a+b}.$$

Alors les équations de la page 114 qui expriment que les φ_{hg} et ψ_{kg} forment un système orthogonal et normal par rapport au premier indice, donnent

$$0 = \sum_g \varphi_{1g}\psi_{2g} = \sum_g \varphi_{2g}\psi_{1g},$$

$$1 = \sum_g \varphi_{1g}\psi_{1g} = \sum_g \varphi_{2g}\psi_{2g}$$

On a donc

$$0 = P_2 + \beta_1\gamma_2 = P_1 + \beta_2\gamma_1,$$
$$1 = P_1 + \beta_1\gamma_1 = P_2 + \beta_2\gamma_2$$

On peut prendre

$$\beta_1\gamma_1 = P_2, \qquad \beta_1\gamma_2 = -P_2, \qquad \beta_2\gamma_1 = -P_1, \qquad \beta_2\gamma_2 = P_1,$$

d'où

$$P_{11}^{(n)} = \frac{b + a(1-a-b)^n}{(a+b)}, \qquad P_{21}^{(n)} = \frac{b}{a+b}[1-(1-a-b)^n],$$

$$P_{12}^{(n)} = \frac{a}{(a+b)}[1-(1-a-b)^n], \qquad P_{22}^{(n)} = \frac{a + b(1-a-b)^n}{a+b}.$$

Dans le cas où $a + b = 1$, on aurait $s_2 = 0$ et l'on vérifie directement que dans ce cas, ces formules restent exactes, mais se simplifient sous la forme

$$P_{11}^{(n)} = P_{21}^{(n)} = b = p_{11} = p_{21}, \qquad P_{12}^{(n)} = P_{22}^{(n)} = a = p_{12} = p_{22}.$$

Exemple du cas oscillant. — Comme le montre la discussion précédente, les cas non réguliers ne peuvent se présenter dans l'exemple que nous venons de traiter, où $r = 2$, que si tout se passe comme si le hasard n'intervenait pas. En effet, dans le cas V les probabilités sont les mêmes que si aucun changement d'état n'était possible; et dans le cas IV les mêmes que si chaque épreuve amenait nécessairement un changement de l'état E_1 à l'état E_2 ou de l'état E_2 à l'état E_1.

Il est donc intéressant de considérer des cas où $r > 2$ pour donner des exemples de cas non réguliers où le hasard intervient effectivement. Nous avons donné des exemples du cas où les $P_{hk}^{(n)}$ ont des limites dépendant de h au cours de l'étude du battage des cartes. Donnons un exemple des cas où les $P_{jk}^{(n)}$ n'ont pas tous des limites déterminées.

Supposons qu'il y ait trois états possibles E_1, E_2, E_3, que de E_2, on passe forcément à E_3 ou de E_3 à E_2, enfin que de E_1 on passe

forcément, soit à E_1 soit à E_2, mais avec des probabilités égales $\left(\text{donc égales à } \frac{1}{2}\right)$. Alors D devient

$$\left\| \begin{matrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & 1 & 0 \end{matrix} \right\|$$

On voit alors facilement qu'on passe de $D^{(n)}$ à $D^{(n+1)}$ en permutant simplement la deuxième et la troisième colonne de $D^{(n)}$. Les $P_{jk}^{(n)}$ seront constants, sauf si j et k sont > 1, cas où ils seront égaux alternativement à 1 et 0. Nous allons donner des exemples algébriquement moins simples, mais ayant une signification concrète intéressante.

Retour au mouvement circulaire — Revenons à l'exemple de la page 98 On a trouvé que si le nombre des positions sur le cercle est impair, il suffit que les probabilités de passage en une épreuve d'une position à une position voisine soient toutes $\neq 0$ pour qu'on soit dans le cas positivement régulier

Nous allons maintenant nous affranchir de cette hypothèse et pour simplifier les calculs supposer le nombre r des positions égal à trois Alors l'équation en s, est, avec les notations de la page 100,

$$0 = \begin{vmatrix} -s & q_2 & p_3 \\ p_1 & -s & q_3 \\ q_1 & p_2 & -s \end{vmatrix},$$

ou en ajoutant les lignes et mettant $(1-s)$ en facteur

$$0 = (1-s) \begin{vmatrix} -s & q_2 & p_3 \\ p_1 & -s & q_3 \\ 1 & 1 & 1 \end{vmatrix},$$

ou

$$(1-s)\{s^2 + s + p_1 p_3 + q_2 q_3 - p_1 q_2\} = 0,$$
$$(1-s)[s^2 + s + p_1 p_2 p_3 + q_1 q_2 q_3] = 0.$$

Il est clair que, pour $s = 1$, le crochet est ≥ 2, donc 1 ne peut être racine multiple, on est certainement dans le cas semi-régulier.

S'il y a une racine $\neq 1$ et de module 1, ou bien elle est réelle, c'est $s = -1$ et alors

$$(16) \qquad p_1 p_2 p_3 + q_1 q_2 q_3 = 0,$$

ou bien elle est complexe, alors il y a une autre racine complexe conjuguée

donc de module 1, et alors

$$p_1 p_2 p_3 + q_1 q_2 q_3 = 1 \tag{17}$$

Pour avoir (16) il faut que $p_1 p_2 p_3 = 0$ et $q_1 q_2 q_3 = 0$, il faut donc que sur les trois probabilités p_1, p_2, p_3 il y en ait au moins une égale à zéro et au moins une autre égale à 1

L'équation (17) peut s'écrire

$$p_1(1 - p_2 p_3) + q_1(1 - q_2 q_3) = 0.$$

Donc $p_1(1 - p_2 p_3)$ et $q_1(1 - q_2 q_3)$ qui sont $\geqq 0$ doivent être nuls Alors $1 - p_2 p_3$ et $1 - q_2 q_3$ ne peuvent être tous deux $\neq 0$. Si par exemple $p_2 p_3 = 1$, on doit avoir $p_2 = p_3 = 1$ et alors $q_1 = 0$, c'est-à-dire $p_1 = 1$ Si $q_2 q_3 = 1$, on aura $q_2 = q_3 = 1$ et alors $p_1 = 0$, c'est-à-dire $q_1 = 1$.

Ainsi : ou bien p_1, p_2, p_3 sont tous trois nuls ou tous trois égaux à 1 et la période asymptotique de $P_{hk}^{(n)}$ est égale à 3; ou bien sur les trois probabilités p_1, p_2, p_3 il y en a au moins deux égales, l'une à 0, l'autre à 1 —, et alors il en sera de même de q_1, q_2, q_3 — et la période asymptotique sera égale à 2

Dans le premier cas, tout se passe comme si le hasard n'intervenait pas presque sûrement le mobile se déplace à chaque épreuve d'un rang dans un sens déterminé. Dans le second cas · si par exemple $p_2 = 0$, $p_3 = 1$, on a $p_{23} = 0$, $p_{31} = 0$, c'est-à-dire que presque certainement le mobile sera au point A_1 de deux en deux épreuves. Et même, dans ce cas les $P_{jk}^{(n)}$ sont exactement de période égale à 2

On voit qu'au contraire, on sera dans le cas régulier dans les autres cas, de sorte que ce cas régulier pourra se présenter même si $p_1, p_2, p_3, q_1, q_2, q_3$ ne sont pas tous $\neq 0$. Si, par exemple, $p_3 = 0$, d'où $q_3 = 1$, on sera dans le cas régulier si l'on n'a

$$\text{ni} \quad p_1 = 1, \quad \text{ni} \quad p_2 = 1, \quad \text{ni} \quad p_1 = p_2 = 0.$$

Alors les expressions (67) de la page 100 sont valables, on a

$$\frac{P_1}{q_2} = \frac{P_2}{1} = \frac{P_3}{p_2 + q_2 q_1} = \frac{1}{2 + q_1 q_2}, \tag{18}$$

et l'on voit qu'on est encore dans le cas positivement régulier.

Deux exemples du cas semi-régulier. — I. *Mouvement circulaire.* — Revenons d'abord au mouvement circulaire précisé page 98, mais en considérant maintenant le cas laissé de côté, celui où le nombre des positions possibles est *pair*. Nous avons vu qu'on est alors dans le cas singulier et, en particulier, qu'on a $P_{jk}^{(n)} = 0$ quand $|j - k|$ et n sont de parités distinctes.

Il est alors naturel d'examiner le comportement asymptotique de $P_{jk}^{(n)}$ quand

n croît par valeurs de même parité. Or, on a

$$P_{jk}^{(n+2)} = \sum_i P_{ji}^{(n)} P_{ik}^{(2)}.$$

En prenant d'abord n pair, $n = 2t$, on pourra donner à j, k, soit des valeurs paires, soit des valeurs impaires.

Posons

$$\varpi_{j',k'}^{(t)} = P_{2j',2k'}^{(2t)}, \qquad \varpi_{j',k'} = P_{2j',2k'}^{(2)} = \varpi_{j',k'}^{(1)},$$

on aura, en posant $i = 2m$,

$$\varpi_{j',k'}^{(t+1)} = \sum_{i'=1}^{i'=m} \varpi_{j'i'}^{(t)} \varpi_{i'k'},$$

$$\varpi_{j',k'} \geqq 0, \qquad \sum_{k'=1}^{k'=m} \varpi_{j'k'} = 1$$

De sorte qu'on se trouve ramené au mouvement d'un point qui ne pourrait occuper que les positions A_2, A_4, .., A_{2m}, chacune des nouvelles épreuves consistant dans l'ensemble de deux consécutives des anciennes épreuves. Or, on a

$$\varpi_{j'k'} = \sum_{i=1}^{i=2m} p_{2j',i} \, p_{i,2k'}$$

D'où

$$\varpi_{j'j'} = p_{2j',2j'-1} \, p_{2j'-1,2j'} + p_{2j',2j'+1} \, p_{2j'+1,2j'} \qquad \text{pour} \quad j \neq m,$$

et

$$\varpi_{mm} = p_{2m,2m-1} \, p_{2m-1,2m} + p_{2m,1} \, p_{1,2m},$$

$$\varpi_{j',j'-1} = p_{2j',2j'-1} \, p_{2j'-1,2j'-2}, \qquad \varpi_{j',j'+1} = p_{2j',2j'+1} \, p_{2j'+1,2j'+2}$$

Si donc on suppose que les $p_{j,j-1}$, $p_{j,j+1}$, p_{1i} et p_{i1} sont tous $\neq 0$, on voit que

$$\varpi_{j'j'} \neq 0, \qquad \varpi_{j',j'-1} \neq 0, \qquad \varpi_{j',j'+1} \neq 0,$$

c'est-à-dire qu'on est dans le cas de Hostinský (p. 33) en ce qui concerne les $\varpi_{j'k'}^{(t)}$. Par conséquent,

$$P_{2j',2k'}^{(2t)} = \varpi_{j',k'}^{(t)}$$

tend vers une limite $\varpi_{k'} \neq 0$ quand t croît indéfiniment. On verrait de même que $P_{2j'-1,2k'-1}^{(2t)}$ tend aussi vers une limite $\varpi'_{k'} \neq 0$ quand t croît indéfiniment. Il en résulte que $P_{jk}^{(n)}$ tend vers une limite déterminée P'_{jk} quand n croît par valeurs paires : quand j et k sont pairs, $P'_{jk} = P'_k = \varpi_{\frac{k}{2}} \neq 0$, quand j et k sont impairs, $P'_{jk} = P'_k = \varpi'_{\frac{k+1}{2}} \neq 0$, quand $|j-k|$ est impair, $P'_{jk} = 0$.

Mais alors $P_{jk}^{(n)}$ tend aussi vers une limite déterminée P''_{jk} quand n croît par

valeurs impaires. Car, on a

$$P_{j,k}^{(2l+1)} = \sum_{i=1}^{i=2m} p_{ji} P_{ik}^{(2l)},$$

donc $P_{jk}^{(2l+1)}$ tend vers

$$P''_{jk} = \sum_{i=1}^{i=2m} p_{ji} P'_{ik}.$$

Ainsi la période asymptotique de $P_{jk}^{(n)}$ est égale à 2. D'ailleurs, on a

$$P''_{jk} = p_{j\,j-1} P'_{j-1,k} + p_{j\,j+1} P'_{j+1,k},$$

sauf

$$P''_{1k} = p_{1,2m} P'_{2m,k} + p_{12} P'_{2k},$$
$$P''_{2m,k} = p_{2m,1} P'_{1k} + p_{2m\,2m-1} P'_{2m-1,k}.$$

Si donc j et k sont de mêmes parités $P''_{jk} = 0$, dans le cas contraire, on a

$$P''_{jk} = p_{j,j-1} P'_k + p_{j,j+1} P'_k = \left(\sum_i p_{ji}\right) P'_k = P'_k$$

Il y a, comme on sait, convergence en moyenne de $P_{jk}^{(n)}$. On aura même, avec des notations évidentes,

$$\Pi_{jk} = \lim_{n \to \infty} \frac{1}{n} [P_{jk}^{(1)} + \ldots + P_{jk}^{(n)}]$$
$$= \lim_{n \to \infty} \left\{ \frac{n'}{n} \left(\frac{P_{jk}^{(1)} + P_{jk}^{(3)} + \ldots + P_{jk}^{(2n'-1)}}{n'} \right) + \frac{n''}{n} \frac{P_{jk}^{(2)} + \ldots + P_{jk}^{(2n'')}}{n''} \right\}$$
$$= \frac{1}{2} (P'_{jk} + P''_{jk})$$

Quand $|j - k|$ est pair,

$$P'_{jk} = P'_k, \qquad P''_{jk} = 0;$$

quand $|j - k|$ est impair

$$P'_{jk} = 0, \qquad P''_{jk} = P'_k,$$

dans tous les cas

$$\Pi_{jk} = \frac{1}{2} P'_k$$

On est donc dans *le cas semi-régulier* : la limite en moyenne arithmétique de $P_{jk}^{(n)}$ est une quantité $\Pi_k = \frac{1}{2} P'_k$ indépendante de l'état initial. Pour déterminer directement les Π_k, il suffit, comme on sait, de résoudre les

équations

$$\Pi_k = \sum_i \Pi_i p_{ik}, \qquad \sum_i \Pi_i = 1,$$

qui deviennent ici

$$\sum_i \Pi_i = 1, \qquad \Pi_k = \Pi_{k-1} p_{k-1\,k} + \Pi_{k+1} p_{k+1,k},$$

sauf

$$\Pi_1 = \Pi_2 p_{21} + \Pi_{2m} p_{2m,1}, \qquad \Pi_{2m} = \Pi_{2m-1} p_{2m-1,2m} + \Pi_1 p_{1,2m}.$$

Par exemple, pour $r = 4$, on trouve, avec les notations de la page 100,

$$(19) \quad \begin{cases} \sum \Pi_k = 1, \qquad \Pi_k = \Pi_{k-1} p_{k-1} + \Pi_{k+1} q_{k+1}, \\ \text{sauf} \\ \Pi_1 = \Pi_4 p_4 + \Pi_2 q_2, \qquad \Pi_4 = \Pi_3 p_3 + \Pi_1 q_1 \end{cases}$$

D'où

$$\frac{\Pi_1}{p_3 p_4 + q_2 q_3} = \frac{\Pi_2}{p_4 p_1 + q_3 q_4} = \frac{\Pi_3}{p_1 p_2 + q_4 q_1} = \frac{\Pi_4}{p_2 p_3 + q_1 q_2}$$
$$= \frac{1}{(p_1 + p_3)(p_2 + p_4) + (q_1 + q_3)(q_2 + q_4)}.$$

Il est visible que si $p_1 = p_2 = p_3 = p_4$ et par suite $q_1 = q_2 = q_3 = q_4$, on aura $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4$. Mais ce résultat peut avoir lieu dans un cas plus général. Si $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4$ $\left(\text{et par suite} = \frac{1}{4}\right)$, les équations (19) deviennent ici :

$$1 = p_4 + q_2, \qquad 1 = p_1 + q_3, \qquad 1 = p_2 + q_4, \qquad 1 = p_3 + q_1,$$

équations dont l'ensemble est équivalent au système

$$p_2 = p_4, \qquad p_1 = p_3.$$

Ainsi la condition nécessaire et suffisante pour que la limite en moyenne de $P_{jk}^{(n)}$ soit indépendante de j et de k est que l'on ait à la fois $p_2 = p_4$, $p_1 = p_3$ c'est-à-dire que p_{jk} ne soit pas modifié quand on ajoute 2 à j et à k (module 4).

II. *Schéma d'urnes.* — MM. Onicescu et Mihoc [1] ont signalé et étudié un intéressant schéma d'urnes qui présente certaines analogies (et aussi certains contrastes) avec celui du mélange des urnes : on considère une seule urne contenant B boules blanches et N boules noires, on procède à une suite de tirages en remplaçant à chaque fois la boule extraite par une boule de l'autre couleur.

Dans les deux schémas, on pourrait, sans calcul, se rendre compte intuitivement de ce qui doit se passer. Dans le mélange d'urnes, supposons que la pro-

portion des blanches soit beaucoup plus forte dans une des urnes, par exemple dans U, que leur proportion $\frac{B}{T}$ dans l'ensemble des urnes. Alors elle sera plus petite dans l'autre urne, de sorte que pendant un certain temps, on tirera probablement plus de blanches de U que de V, et les proportions dans U et V tendront probablement à se rapprocher de $\frac{B}{T}$. C'est ce que notre calcul a confirmé. Dans les schémas de MM Onicescu et Mihoc, si à un moment donné le nombre des boules blanches est sensiblement plus élevé que celui des noires, il y aura de grandes chances de tirer une blanche, qui sera remplacée par une noire et par suite les proportions des boules blanches et noires tendront à s'égaliser C'est aussi ce que confirme les calculs des deux auteurs Toutefois en prenant comme variable, non pas le nombre b des boules blanches figurant dans l'urne après n tirages, mais le nombre β des boules blanches extraites au cours des n tirages, ces auteurs ont pu établir ce résultat intéressant que la fréquence $\frac{\beta}{n}$ ne tend pas seulement vers sa limite comme dans le cas de Bernoulli, c'est-à-dire « en probabilité » (¹) et même « presque sûrement » (¹), mais encore, selon l'expression de ces auteurs, « au sens de l'Analyse classique » C'est-à-dire que quelle que soit la suite d'épreuves qu'on puisse effectuer, celle-ci une fois réalisée, on constatera que la suite des valeurs $\frac{\beta}{n}$ converge sûrement vers $\frac{1}{2}$.

Leur démonstration, donnée en supposant $B = N$, mais qui subsiste sans cette hypothèse, résulte immédiatement du fait qu'on aura $b = B - \beta + n - \beta$, avec $0 \leqq b \leqq B + N$, d'où

$$\frac{-B}{2n} \leqq \frac{1}{2} - \frac{\beta}{n} \leqq \frac{N}{2n},$$

et par suite $\lim\limits_{n \to \infty} \frac{\beta}{n} = \frac{1}{2}$.

Les auteurs ont ensuite cherché à déterminer ce que devient, quand n croît, la probabilité $\Pi_\beta^{(n)}$ pour que le nombre des boules blanches extraites en n épreuves soit égal à β. Ils y sont parvenus par l'intermédiaire de la notion de fonction caractéristique, et en remplaçant la variable β dont la borne supérieure croît indéfiniment avec n, par l'écart $x = \frac{1}{2} - \frac{\beta}{n}$ qui reste borné quand n croît.

On peut y arriver aussi [Fréchet, 195] par la méthode (celle de Markoff) que nous avons appliquée dans le cas du mélange des urnes. Chemin faisant nous préciserons un peu l'interprétation des calculs de ces deux auteurs et nous completerons leurs résultats. Seulement, nous allons voir que, contrairement au cas du mélange des urnes, on n'est pas ici dans le cas régulier, de

(¹) Voir Fréchet [188], p. 164 et 215.

sorte qu'il faut modifier un peu la démonstration. On peut, à cet effet, employer un raisonnement *qui pourrait servir pour d'autres exemples de cas singuliers*, qui nous a servi déjà, page 123, et qui consiste à ramener au cas régulier en considérant comme épreuve unique un groupement convenable des épreuves primitivement considérées.

Appelons encore $P_{hk}^{(n)}$ la probabilité de passer, en n tirages, d'un effectif de h boules blanches dans l'urne à un effectif de k. Comme on a $B + n - b = 2\beta$, on voit que $B - b$ et n doivent être de même parité. De sorte que $P_{Bb}^{(n)} = 0$ dans le cas contraire, c'est-à-dire que si $|B - b|$ est impair, $P_{Bb}^{(2)} = P_{Bb}^{(4)} = \ldots = 0$, et si $|B - b|$ est pair, $P_{Bb}^{(1)} = P_{Bb}^{(3)} = \ldots = 0$. Si $P_{Bb}^{(n)}$ avait une limite unique indépendante de B quand n croît, cette limite serait nulle, quelle que soit b, ce qui est incompatible avec la condition $\sum_b P_{Bb}^{(n)} = 1$. On n'est donc pas dans le cas régulier. Mais ce qui précède conduit à se demander si l'on n'obtiendrait pas une limite unique en faisant croître n par valeurs de même parité, puisque cela a déjà lieu quand n garde une parité distincte de celle de $|B - b|$.

L'artifice dont nous avons parlé consistera alors à considérer chaque couple de tirages de rangs différant de deux unités comme une épreuve unique. Et comme on a

$$P_{Bb}^{(n+2)} = \sum_j P_{Bj}^{(n)} P_{jb}^{(2)},$$

on pourra poser

$$\varpi_{Bb}^{(m)} = P_{bB}^{(2m)}, \qquad \varpi'^{(m)}_{Bb} = P_{Bb}^{(2m+1)}, \qquad \varpi_{jb} = P_{jb}^{(2)}.$$

De sorte que

$$\varpi_{Bb}^{(m+1)} = \sum_j \varpi_{Bj}^{(m)} \varpi_{jb}, \tag{20}$$

avec

$$\varpi_{jb} \geqq 0, \qquad \sum_b \varpi_{jb} = 1.$$

On est ramené à un problème analogue à celui de Markoff où le rôle des p_{jb} est tenu par celui des ϖ_{jb}, et où les valeurs possibles de j et b sont

$$(0 \leqq) \ldots, \quad B - 2, \quad B, \quad B + 2, \quad \ldots (\leqq B + N). \tag{21}$$

Or, il est facile de calculer les ϖ_{jb}. On a

$$p_{jb} = 0 \quad \text{si} \quad |j - b| \neq 1, \qquad p_{b-1,b} = 1 - \frac{b-1}{T}, \qquad p_{b+1,b} = \frac{b+1}{T},$$

d'où $\varpi_{jb} = P_{jb}^{(2)} = 0$, si $|j-b|$ est différent de 0 ou 2, avec

$$\varpi_{b-2,b} = P_{b-2,b}^{(2)} = p_{b-2,b-1}\,p_{b-1,b} = \left(1-\frac{b-2}{T}\right)\left(1-\frac{b-1}{T}\right),$$

$$\varpi_{b+2,b} = p_{b+2,b+1}\,p_{b+1,b} = \frac{(b+2)(b+1)}{T^2},$$

$$\begin{aligned}\varpi_{bb} &= p_{b,b-1}\,p_{b-1,b}+p_{b,b+1}\,p_{b+1,b}\\ &= \frac{b}{T}\left(1-\frac{b-1}{T}\right)+\left(1-\frac{b}{T}\right)\frac{b+1}{T} = \frac{2b+1}{T}-\frac{2b^2}{T^2},\end{aligned}$$

avec toutefois une rectification à faire tenant compte de la disparition du terme $p_{b,b-1}p_{b-1,b}$ si b prend la valeur zéro. De même, disparition de l'autre terme si b prend la valeur de $B+N$.

Dès lors, si nous n'admettons plus pour les ϖ_{jb} que les valeurs de j, b de même parité, nous nous trouvons dans le cas de Hostinský (p. 33). On est donc aussi dans le cas régulier en ce qui concerne les $\varpi_{Bb}^{(n)}$ et même dans le cas positivement régulier. Finalement nous avons démontré que pour $|B-b|$ pair $P_{Bb}^{(2m)}$ tend vers une limite positive ϖ_b indépendante de B. D'ailleurs, pour $|B-b|$ impair, $P_{Bb}^{(2m)}$, restant nul, tend vers une limite qui est zéro. On trouve ϖ_b en passant à la limite dans (20), ce qui donne

$$\varpi_b = \sum_j \varpi_j\varpi_{jb} = \varpi_{b-2}\varpi_{b-2,b}+\varpi_b\varpi_{bb}+\varpi_{b+2}\varpi_{b+2,b}, \tag{22}$$

c'est-à-dire

$$\begin{aligned}\varpi_b = {}&\varpi_{b-2}\left(1-\frac{b-2}{T}\right)\left(1-\frac{b-1}{T}\right)\\ &+\varpi_b\left[\frac{b}{T}\left(1-\frac{b-1}{T}\right)+\left(1-\frac{b}{T}\right)\frac{b+1}{T}\right]+\varpi_{b+2}\frac{(b+2)(b+1)}{T^2},\end{aligned} \tag{23}$$

avec toutefois disparition au second membre du premier produit si $b<2$ et du dernier si $b>B+N-2$. Soit b_0 la plus petite valeur de b ($b_0=0$ si B est pair, $b_0=1$ si B est impair). On aura

$$\varpi_{b_0} = \varpi_{b_0}\varpi_{b_0,b_0}+\varpi_{b_0+2}\varpi_{b_0+2,b_0},$$

qui permettra de calculer ϖ_{b_0+2} supposant connu ϖ_{b_0}, en remplaçant ensuite successivement b par b_0+2, b_0+4, b_0+6, ... dans (23), on en déduira successivement ϖ_{b_0+4}, ϖ_{b_0+6}, ... supposant connu ϖ_{b_0}. Enfin, l'égalité $\Sigma\varpi_b=1$ permettra de calculer ϖ_{b_0} et par suite ϖ_b quel que soit b. (C'est la méthode déjà employée pour le mélange des urnes, page 49.)

Supposons, par exemple, pour commencer, B et N pairs. Alors $b_0=0$, et l'on a

$$\varpi_0 = \varpi_0\varpi_{00}+\varpi_2\varpi_{20},$$

avec

$$\varpi_{00} = \frac{1}{T}, \qquad \varpi_{20} = \frac{2}{T^2},$$

d'où

$$\varpi_2 = \varpi_0 \frac{T(T-1)}{2};$$

puis $\varpi_4 = \varpi_0 \varpi_{02} + \varpi_2 \varpi_{22} + \varpi_4 \varpi_{42}$, avec

$$\varpi_{22} = \frac{5}{T} - \frac{8}{T^2}, \qquad \varpi_{42} = \frac{4.3}{T^2}, \qquad \varpi_{02} = 1\,\frac{1}{T},$$

d'où

$$\varpi_4 = \varpi_0 \frac{T(T-1)(T-2)(T-3)}{1.2.3.4}.$$

Nous voyons qu'on a pour $\alpha = 0, 2, 4$,

$$\varpi_\alpha = \varpi_0 C_T^\alpha \tag{24}$$

On va voir que cette dernière relation est générale. En effet, comme elle garde son sens ainsi que (23) sans supposer B et N pairs, on peut abandonner cette dernière hypothèse et se contenter, puisque la solution, s'il en existe une, est unique, de vérifier, ce qui n'offre pas de difficultés, que la relation (23) est satisfaite quand on y fait $\varpi_\alpha = \varpi_0 C_T^\alpha$ pour $\alpha = b-2, b, b+2$.

Ainsi, nous avons prouvé qu'on a

$$\varpi_b = \varpi_0 C_T^b,$$

quand b prend les valeurs (21).

Il ne reste plus qu'à trouver ϖ_0, ce qui résulte de l'égalité

$$1 = \sum_0 \varpi_b = \varpi_0 \sum_0 C_T^b.$$

Ici $\sum_0 C_T^b$ est la somme de ceux des termes $C_T^0 + C_T^1 + C_T^2 + \ldots + C_T^{T-1} + C_T^T$ (où l'on a posé $C_T^0 = 1 = C_T^T$) dont les indices sont de même parité que B. C'est donc l'une des deux sommes $1 + C_T^2 + C_T^4 + \ldots$, $C_T^1 + C_T^3 + C_T^5 + \ldots$ qui sont, puisque $(1-1)^T = 0$, égales entre elles et à 2^{T-1}. D'où $\varpi_0 = \frac{1}{2^{T-1}}$, et en général

$$\varpi_b = \frac{1}{2^{T-1}} C_T^b.$$

En résumé, *la loi de probabilité exprimée par l'expression de* $P_{Bb}^{(n)}$ *en fonction de b tend, quand n croît par valeurs paires, vers une limite* ϖ_b *égale à zéro pour les valeurs de b de parité différente de celle de* B *et,*

pour les autres valeurs de b, égales à $\frac{1}{2^{T-1}} C_T^b$. On verrait de même que $P_{Bb}^{(n)}$ tend, quand n croît *par valeurs impaires*, vers une limite ϖ'_b égale à zéro pour les valeurs de b de même parité que B et vers $\frac{1}{2^{T-1}} C_T^b$ pour les autres valeurs de b.

Nous venons de vérifier sur cet exemple particulier que $P_{Bb}^{(n)}$ est une fonction asymptotiquement périodique de n, *de période asymptotique égale à* 2, cette fonction de b converge en moyenne arithmétique, sa limite Π_b en moyenne arithmétique étant évidemment $\frac{1}{2}(\varpi_b + \varpi'_b)$. Comme l'une de ces quantités est égale à 0 et l'autre à $\frac{1}{2^{T-1}} C_T^b$, on voit que *la limite en moyenne arithmétique*, Π_b *de* $P_{Bb}^{(n)}$, *est la probabilité binomiale*, $\Pi_b = \frac{1}{2^T} C_T^b$, *c'est-à-dire de la probabilité de la répétition* b *d'un événement de probabilité* $\frac{1}{2}$ *en* T *épreuves*.

Il en résulte immédiatement que si le nombre de boules T est grand, la loi limite en moyenne « réduite » est approximativement la loi de Gauss, la valeur moyenne de b étant $\frac{T}{2}$ et son écart quadratique moyen $\frac{\sqrt{T}}{2}$.

D'ailleurs, on peut étendre ces deux derniers résultats au cas où T n'est pas supposé très grand. En effet, la valeur moyenne M_n de b est alors

$$M_n = \sum_b b P_{Bb}^{(n)},$$

elle tend quand n croît par valeurs de même parité vers

$$M = \sum_b{}' b \frac{C_T^b}{2^{T-1}},$$

où Σ' est étendu à des valeurs de b de même parité et qui sont ≥ 0, $\leq T$. La valeur de M est la même pour chacune des deux parités.

En effet, on a

$$M = \sum_b{}' \frac{T C_{T-1}^{b-1}}{2^{T-1}} = T \frac{\sum_b{}' C_{T-1}^{b-1}}{2^{T-1}}.$$

Or $\sum_b{}' C_{T-1}^{b-1} = 2^{T-2}$ quelle que soit la parité de b, on a donc . $M = \frac{T}{2}$. Ainsi se trouve justifiée l'affirmation déduite tout d'abord d'un raisonnement intuitif : *la valeur moyenne du nombre de boules blanches contenues dans l'urne après n tirages tend, quand n croît en passant par toutes les*

valeurs entières, vers une limite unique qui est la moitié du nombre total des boules de l'urne.

D'autre part, l'écart quadratique moyen μ_n de b est donné par

$$\mu_n^2 = \sum_b (b - M_n)^2 P_{Bb}^{(n)}.$$

Quand n croît par valeurs de même parité, μ_n^2 tend vers

$$\mu^2 = \sum_b{}' (b - M)^2 \frac{C_T^b}{2^{T-1}}.$$

Un calcul analogue à celui de la page 75 fournit

$$\mu^2 = \sum_b{}' b(b-1)\frac{C_T^b}{2^{T-1}} + M - M^2$$

$$= T(T-1)\sum_{b \geqq 2}{}' \frac{C_{T-2}^{b-2}}{2^{T-1}} + \frac{T}{2} - \frac{T^2}{4} = T(T-1)\frac{2^{T-3}}{2^{T-1}} + \frac{T}{2} - \frac{T^2}{4} = \frac{T}{4}.$$

On voit que la limite est ici encore indépendante de la parité des valeurs de n considérées. *L'écart quadratique moyen μ_n du nombre de boules blanches contenues dans l'urne après n tirages converge quand n croît par valeurs entières consecutives vers une limite unique $\mu = \frac{\sqrt{T}}{2}$.*

Une fois calculés les Π_k, $\left(\Pi_k = \frac{C_T^k}{2^T}\right)$, on aura les deux limites P'_{jk}, P''_{jk}, des $P_{jk}^{(n)}$, par les règles

		Et si n croît par valeurs paires	impaires
$P_{jk}^{(n)}$ tend	si $\vert j - k \vert$ est pair	vers $2\Pi_k$	vers zéro
	si $\vert j - k \vert$ est impair	vers zéro	vers $2\Pi_k$

On voit que les limites P'_{jk}, P''_{jk}, sans être indépendantes de j, ne dépendent que de sa parité.

MM. Onicescu et Mihoc avaient supposé $B = N$. Si nous avons supprimé cette hypothèse, c'est parce que son abandon nous a permis, comme on le voit maintenant, d'établir que les limites M, μ de la moyenne M_n et de la dispersion μ_n de la composition de l'urne (composition définie par la valeur de b), sont indépendantes de la composition initiale.

b — Variables en chaîne.

Un certain nombre des résultats concernant les variables en chaîne dans le cas régulier (p. 67-102) peuvent être étendus au cas général (Fréchet [147], p. 154-164), grâce à l'introduction de la convergence en moyenne arithmétique des $P_{lk}^{(n)}$.

Valeurs moyennes. — Reprenons les notations de la page 67; on voit d'abord, d'après la formule (37), que *la valeur moyenne* $\mathfrak{M}X_h^{(n)}$ *du nombre aléatoire* X(E) *après n épreuves converge toujours au moins au sens de Cesaro* (déf. p. 259) *quand n croît*.

On peut exprimer autrement ce même résultat en considérant la valeur moyenne $M_h^{(n)}$ au sens des probabilités, de la moyenne arithmétique $A_h^{(n)}$ des valeurs prises par X(E) dans les *n* premières épreuves. On voit alors que *cette valeur moyenne* $M_h^{(n)}$ *converge toujours* (*au sens ordinaire*) *quand n croît*. On a ainsi étendu au cas général une proposition démontrée par Markoff dans le cas régulier et étendue par M. von Mises au cas où le déterminant D des p_{ik} est « indécomposable » (p. 168).

La limite M_h au sens de Cesaro de $\mathfrak{M}X_h^{(n)}$ et la limite au sens ordinaire de $M_h^{(n)}$ sont évidemment égales entre elles et à $\sum_l \Pi_{hl} x_l$, elles peuvent dépendre de l'état initial. En vertu de (9), page 110, on peut aussi écrire

$$M_h = \sum_g U_{hg} T_g,$$

où les U_{hg} sont indépendants des valeurs possibles $x_1, x_2, \ldots$ de X(E), et où

$$T_g = \sum_l V_{lg} x_l$$

Pour que cette limite M_h soit indépendante de l'état initial, quelles que soient les valeurs possibles $x_1, x_2, \ldots x_l$ de Y(E), il faut et il suffit qu'on soit dans le cas semi-régulier, c'est-à-dire que l'unité soit racine simple de l'équation en s.

La quantité $M_h^{(n)} - M_h$ est infiniment petite avec $\frac{1}{n}$; on peut même

ajouter qu'*elle est au moins de l'ordre de* $\frac{1}{n}$. Plus précisément, le produit $n(M_h^{(n)} - M_h)$ est borné quand n croît, puisque

$$n(M_h^{(n)} - M_h) = \sum_l n(\Pi_{hl}^{(n)} - \Pi_{hl})\, x_l,$$

et que $n(\Pi_{hl}^{(n)} - \Pi_{hl})$ est borné. Dans le cas, régulier ou non, où les $P_{hl}^{(n)}$ convergent au sens ordinaire, le produit $n(M_h^{(n)} - M_h)$ converge vers $\sum_l s_{hl} x_l$ qui est la somme de la série absolument convergente

$$\sum_{t=1}^{t=+\infty} \left\{ \sum_{l=1}^{l=r} (P_{hl}^{(t)} - P_{hl})\, x_l \right\}.$$

Dans le cas général, la valeur moyenne $\mathfrak{M}Y_h^{(n)}$ est *une fonction asymptotiquement périodique* de n.

Les racines de $\Delta(s)$ qui sont de modules 1 étant racines d'une même équation binome $s^N = 1$, N est une période asymptotique de $\mathfrak{M}Y_h^{(n)}$. Pour que $\mathfrak{M}Y_h^{(n)}$ soit une fonction périodique de n, quelles que soient les valeurs possibles de X(E), il faut et il suffit que $\Delta(s)$ n'ait pas de racine de module $\neq 1$. Pour que $\mathfrak{M}Y_h^{(n)}$ converge au sens ordinaire quand n croît quelles que soient les valeurs possibles de X(E), il faut et il suffit que $\Delta(s)$ n'ait pas de racine de module 1 autre que 1.

On peut donner une sorte de développement limité de $M_l^{(n)}$. On sait, en effet (Note B, p. 276), que

$$n(\Pi_{hl}^{(n)} - \Pi_{hl}) = s_{hl} + U_{hl}(n),$$

où $U_{1l}(n), \ldots, U_{rl}(n)$ forment un système de solutions, convergeant en moyenne vers zéro avec $\frac{1}{n}$ — de

$$\text{(H)} \qquad u_j(n) - \sum_k p_{jk}\, u_k(n-1) = 0$$

Donc

$$(24\ bis) \qquad n(M_h^{(n)} - M_h) = \sum_l s_{hl} x_l + \varphi_j^{(n)}$$

où

$$\varphi_j^{(n)} = \sum_l U_{jl}(n)\, x_l,$$

est aussi une solution convergeant en moyenne vers zéro du système (H). De sorte que

$$\varphi_j^{(n)} = \xi_j(n) + \lambda_j(n), \tag{24 ter}$$

où $\xi_j(n)$ converge exponentiellement vers zéro et $\lambda_j(n)$ est une fonction périodique de n, convergeant en moyenne vers zéro.

Dispersion. — On peut conserver dans le cas général (Fréchet, [**147**], p. 154-164) quelques-uns des résultats concernant la dispersion obtenus par Markoff dans le cas régulier. Au lieu du raisonnement appliqué directement au cas général dans notre mémoire, nous pouvons nous contenter de voir ici ce qui subsiste, dans le cas général, des raisonnements faits dans le cas régulier page 74-80, en ce qui concerne le comportement de $\mu_h^{(n)}$, $\lambda_h^{(n)}$, $\rho_h^{(n)}$, $\delta_h^{(n)}$, $\mathcal{G}_{hk}^{(n)}$, $\mathcal{E}_{hk}^{(n)}$

On a d'abord

$$[\lambda_h^{(n)}]^2 = \sum_k (x_k - M_h)^2 P_{hk}^{(n)},$$

d'où

$$\frac{1}{n}\sum_{l=1}^{l=n} [\lambda_h^{(l)}]^2 = \sum_k (x_k - M_h)^2 H_{hk}^{(n)}.$$

Donc $[\lambda_h^{(n)}]^2$ converge toujours au moins au sens de Cesaro quand n croît. Sa limite est

$$[\lambda_h]^2 = \sum_k (x_k - M_h)^2 H_{hk} \tag{25}$$

Dans le cas où les $P_{hk}^{(n)}$ convergent au sens ordinaire, on déduit de la formule

$$[\lambda_h^{(n)}]^2 = [\mu_h^{(n)}]^2 + [M_h - \mathfrak{M}\,Y_h^{(n)}]^2,$$

que $[\mu_h^{(n)}]^2$ et $[\lambda_h^{(n)}]^2$ convergent vers la même limite $[\lambda_h]^2$.

On va maintenant employer les formules (46), (48 *bis*), (49), (50) des pages 76, 77, en y remplaçant M par M_h. $\delta_h^{(n)}$, $\rho_h^{(n)}$ sont donc les écarts quadratiques respectifs moyens de $A_h^{(n)}$ avec M_h, $M_h^{(n)}$, et l'on a

$$n[\rho_h^{(n)}]^2 = n[\delta_h^{(n)}]^2 - \frac{\{n[M_h^{(n)} - M_h]\}^2}{n}.$$

On a vu que $n[M_h^{(n)} - M_h]$ reste toujours borné. Donc

$$n[\rho_h^{(n)}]^2 - n[\delta_h^{(n)}]^2$$

tend vers zéro dans le cas singulier comme dans le cas régulier

Dans la formule (48), puisque $[\lambda_h^{(n)}]^2$ tend vers $[\lambda_h]^2$ au sens de Cesaro, le premier terme tend au sens ordinaire vers $[\lambda_h]^2$, de sorte qu'il reste encore à étudier le comportement de L_n qui est ici

$$L_n = \frac{2}{n} \sum_{u \neq v} \mathfrak{M} \{ (X_h^{(u)} - M_h)(X_h^{(v)} - M_h) \} = 2 \sum_i \sum_j G_{i,j}^{(n)} (x_i - M_h)(x_j - M_h)$$

L'expression (50) de $G_{ij}^{(n)}$ n'est pas à changer. Par contre, pour utiliser $S_{ij}^{(n)}$, il faudra modifier (51) et poser $\varphi_{ij}^{(u)} = P_{ij}^{(u)} - \Pi_{ij}$ On a

$$(25\ bis) \qquad n[G_{ij}^{(n)} - S_{ij}^{(n)}] = \Pi_{ij}[(n-1)P_{hi}^{(1)} + (n-2)P_{hi}^{(2)} + \ldots + P_{hi}^{n-1}]$$
$$= \Pi_{ij} \{ (n-1)\Pi_{hi}^{(n-1)} + \ldots + \Pi_{hi}^{(1)} \}$$
$$= \Pi_{ij} \sum_{t=1}^{t=n-1} t[\Pi_{hi}^{(t)} - \Pi_{hi}] + \frac{n(n-1)}{2} \Pi_{ij}\Pi_{hi}$$

Or, on a vu que $|t(\Pi_{hi}^{t} - \Pi_{hi})|$ a, quand h, i, t varient, une borne supérieure H. Donc

$$\left| \frac{G_{ij}^{(n)} - S_{ij}^{(n)}}{n} - \frac{1}{2}\Pi_{ij}\Pi_{hi} \right| \leqq \frac{\Pi_{ij} H}{n} + \frac{\Pi_{hi}\Pi_{ij}}{2n}.$$

Dès lors, $\frac{G_{ij}^{(n)} - S_{ij}^{(n)}}{n}$ tend vers $\frac{\Pi_{hi}\Pi_{ij}}{2}$ et, plus précisément,

$$\left\{ G_{ij}^{(n)} - S_{ij}^{(n)} - \frac{n}{2}\Pi_{hi}\Pi_{ij} \right\}$$

reste borné quand n croît

Reste donc à étudier $S_{ij}^{(n)}$. On a, dans le cas général,

$$(26) \qquad n S_{ij}^{(n)} = P_{hi}^{(1)}[\varphi_{ij}^{(1)} + \ldots + \varphi_{ij}^{(n-1)}] + \ldots + P_{hi}^{(n-1)}\varphi_{ij}^{(1)}$$
$$= \sum_{t=1}^{t=n-1} P_{hi}^{(t)}(n-t)(\Pi_{ij}^{(n-t)} - \Pi_{ij}),$$

d'où

$$|n S_{ij}^{(n)}| \leqq (n-1) H, \qquad |S_{ij}^{(n)}| \leqq H.$$

Ainsi $S_{ij}^{(n)}$ reste borné.

En résumé $\left\{ G_{ij}^{(n)} - \frac{n}{2} \Pi_{hi}\Pi_{ij} \right\}$ reste toujours borné.

Il en résulte qu'en posant

$$(27) \qquad [W_h]^2 = \sum_i \sum_j \Pi_{hi}\Pi_{ij}(x_i - M_h)(x_j - M_h),$$

d'où

$$(28) \qquad (W_h)^2 = \sum_i \Pi_{hi}(x_i - M_h)(M_i - M_h),$$

la quantité $L_n - n(W_h)^2$ reste aussi bornée. Par suite l'expression

$$T_n = n \{ [\delta_h^{(n)}]^2 - [W_h]^2 \},$$

qui peut s'écrire, d'après (48), page 77

$$T_n = \frac{1}{n} \sum_{u=1}^{u=n} [\lambda_h^{(u)}]^2 + L_n - n(W_h)^2$$

est bornée quand n varie; dès lors, dans le cas singulier comme dans le cas régulier, *l'écart quadratique moyen $\delta_h^{(n)}$ de $X_h^{(n)}$ avec M_h a toujours une limite déterminée* W_h

[Puisque le second membre de (27) est la limite de $[\delta_h^{(n)}]^2$, nous avions bien le droit de le représenter par un carré : $(W_h)^2$.]

D'après (28), la quantité W_h peut être nulle, par exemple quand les $M_j = \sum_i \Pi_{ji} x_i$ sont égaux, et par suite égaux à un nombre M indépendant de j. Pour qu'il en soit ainsi quels que soient les x_i, il faut que dans les relations $\sum_i (\Pi_{ji} - \Pi_{ki}) x_i = 0$, les coefficients des x_i soient nuls. Cette propriété ne peut donc avoir lieu quels que soient les x_i que dans le cas semi-régulier, et elle a certainement lieu dans ce cas. Mais elle peut avoir lieu en dehors du cas semi-régulier, quand les p_{ik} sont donnés arbitrairement, pour des choix convenables des x_i, par exemple quand les x_i sont égaux entre eux. Nous pourrons caractériser le cas où les M_j sont égaux en disant qu'alors *la variable aléatoire* X(E) *est semi-régulière*.

Lorsque W_h est nul, donc, en particulier, *dans le cas semi-régulier*, et même dans le cas d'une variable aléatoire semi-régulière,

non seulement $\delta_h^{(n)}$ tend vers zéro avec $\frac{1}{n}$, mais encore $\delta_h^{(n)}$ *est un infiniment petit d'ordre au moins égal à celui de* $\frac{1}{\sqrt{n}}$.

Cependant W_h peut être effectivement $\neq 0$ (dans des cas nécessairement singuliers), comme nous en verrons plus loin (p 142) des exemples.

Sans supposer encore que W_h est nul, nous allons chercher un développement limité de $[\delta_h^{(n)}]^2$ du premier ordre en $\frac{1}{n}$, lequel nous permettra de préciser l'ordre de $\delta_h^{(n)}$ dans le cas important où $W_h = 0$.

Cas non oscillant. — On a vu que $T_n = n\{[\delta_h^{(n)}]^2 - (W_h)^2\}$ est toujours borné. Il est naturel de se demander s'il y a des cas où cette quantité a une limite.

On voit facilement que T_n a une limite dans le cas non oscillant. Car, dans ce cas, $t[\Pi_{hi}^{(t)} - \Pi_{hi}]$ converge vers s_{hi}, la formule (25 *bis*) donne donc

$$(29) \qquad \lim_{n\to\infty}\left[G_{ij}^{(n)} - S_{ij}^{(n)} - \frac{n-1}{2}\Pi_{hi}\Pi_{ij}\right]$$

$$= \Pi_{ij}\lim_{n\to\infty}\frac{n-1}{n}\left[\frac{\sum_{t=1}^{t=n-1} t(\Pi_{hi}^{(t)} - \Pi_{hi})}{n-1}\right] = s_{hi}\Pi_{ij}$$

Quant à

$$S_{ij}^{(n)} = \frac{n-1}{n}\,\frac{t_1 + \ldots + t_{n-1}}{n-1},$$

avec

$$t_n = P_{hi}^{(1)}\varphi_{ij}^{(n)} + \ldots + P_{hi}^{(n)}\varphi_{ij}^{(1)},$$

on observe que $P_{hi}^{(n)} \to P_{hi}$ et que $\varphi_{ij}^{(1)} + \ldots + \varphi_{ij}^{(n)}$ converge absolument vers s_{ij}, de sorte que [*voir* note C, page 280, premier Lemme] t_n tend vers $P_{hi}s_{ij}$ et que, par suite,

$$\lim_{n\to\infty} S_{ij}^{(n)} = P_{hi}s_{ij}.$$

Des lors

$$\lim_{n\to\infty}\left[G_{ij}^{(n)} - \frac{n}{2}P_{hi}P_{ij}\right] = s_{hi}P_{ij} - \frac{1}{2}P_{hi}P_{ij} + P_{hi}s_{ij},$$

et, par suite,

$$(30)\quad \lim_{n\to\infty} n\{[\delta_h^{(n)}]^2-[W_h]^2\}$$
$$=\sum_i (x_i-M_h)^2 P_{hi}$$
$$+2\sum_i\sum_j\left(s_{hi}P_{ij}+P_{hi}s_{ij}-\frac{1}{2}P_{hi}P_{ij}\right)(x_i-M_h)(x_j-M_h)$$

Cas général. Premières proprietés — Depuis que nous avons etabli l'existence d'une limite de T_n dans le cas non oscillant, M. Dœblin a pu aussi établir l'existence d'une limite dans un autre cas Avant d'y arriver par une nouvelle méthode (p. 152), indiquons d'abord le résultat moins précis fourni par la méthode actuelle

Dans le cas général, on peut écrire

$$S_{ij}^{(n)}=\frac{1}{n}\{P_{hi}^{(1)}[\varphi_{ij}^{(1)}+\ldots+\varphi_{ij}^{(n)}]+\ldots+P_{hi}^{(n)}\varphi_{ij}^{(1)}\}$$
$$=\frac{1}{n}\{p_{hi}U_{ij}^{(n)}+\ldots+P_{hi}^{(n)}U_{ij}^{(1)}\}+\frac{1}{n}[p_{hi}+\ldots+P_{hi}^{(n)}]s_{ij}$$

Le dernier terme converge évidemment vers $\Pi_{hi}s_{ij}$. D'autre part, on a vu (p. 133) que les quantités

$$U_{ij}^{(n)}=\varphi_{ij}^{(1)}+\ldots+\varphi_{ij}^{(n)}-s_{ij}$$

convergent en moyenne vers zéro et forment pour j fixe, un système de solutions de (H).

Puisque $P_{hi}^{(n)}$ et $U_{ij}^{(n)}$ sont des fonctions asymptotiquement périodiques de période N, le premier terme du second membre de $S_{ij}^{(n)}$ est aussi asymptotiquement périodique, d'après le cinquième Lemme de la Note C, page 282, et tend en moyenne vers le produit des limites en moyenne, Π_{hi} et zéro, de $P_{hi}^{(n)}$ et $U_{ij}^{(n)}$, c'est-à-dire vers zéro.

Dès lors

$$\lim_{n\to\infty}\text{moy.}\left\{G_{ij}^{(n)}-\frac{n}{2}\Pi_{hi}\Pi_{ij}\right\}=s_{hi}\Pi_{ij}+\Pi_{hi}s_{ij}-\frac{1}{2}\Pi_{hi}\Pi_{ij},$$

et, par suite, on a

$$\lim_{n\to\infty}\text{moy.}\left[n\{[\delta_h^{(n)}]^2-[W_h]^2\}\right]$$
$$=\sum_i(x_i-M_h)^2\Pi_{hi}+2\sum_i\sum_j(\Pi_{hi}s_{ij}+s_{hi}\Pi_{ij})(x_i-M_h)(x_j-M_h)-[W_h]^2,$$

ce qui donne une sorte de développement limité du premier ordre de $[\delta_h^{(n)}]^2$:

$$(31)\quad [\delta_h^{(n)}]^2 = [W_h]^2$$
$$+\frac{1}{n}\left\{\sum_i (x_i - M_h)^2 \Pi_{hi} + 2\sum_i\sum_j (\Pi_{hi}s_{ij} + s_{hi}\Pi_{ij})(x_i - M_h)(x_j - M_h) - [W_h]^2 + \varepsilon_h(n)\right\},$$

avec

$$\lim_{n\to\infty} \text{moy.}\, \varepsilon_h(n) = 0,$$

et

$$[W_h]^2 = \sum_i\sum_j \Pi_{hi}\Pi_{ij}(x_i - M_h)(x_j - M_h) = \sum_i \Pi_{hi}(x_i - M_h)(M_i - M_h).$$

L'expression de $[\rho_h^{(n)}]^2$ sera exactement la même en remplaçant cependant $\varepsilon_h(n)$ par une autre quantité $\omega_h(n)$, qui converge aussi en moyenne vers zéro

Dans le cas particulier où $W_h = 0$, l'expression de $\rho_h^{(n)}$ se simplifie et devient

$$(32)\quad [\rho_h^{(n)}]^2 = \frac{(\sigma_h)^2 + \omega_h(n)}{n},$$

avec (¹)

$$(33)\quad (\sigma_h)^2 = \sum_i (x_i - M_h)^2 \Pi_{hi} + 2\sum_i\sum_j (\Pi_{hi}s_{ij} + s_{hi}\Pi_{ij})(x_i - M_h)(x_j - M_h)$$

Dans cette expression figure

$$2\sum_i\sum_j s_{hi}\Pi_{ij}(x_i - M_h)(x_j - M_h) = 2\sum_i s_{hi}(x_i - M_h)(M_i - M_h)$$

Dans le cas où les M_i sont égaux à une même quantité M, W_h est nécessairement nul ainsi que cette dernière expression. Donc : dans le cas d'une variable aléatoire semi-régulière, on a :

$$(34)\quad \lim_{n\to\infty} \text{moy}\; n[\rho_h^{(n)}]^2 = (\sigma_h)^2,$$

(¹) Un calcul analogue à celui de la page 84 fournirait la décomposition en carrés suivante (qu'on pourra se contenter de vérifier) .

$$(\sigma_k)^2 = \sum_j \Pi_{kj} \sum_i p_{ji}(z_i + \theta_i - \theta_j)^2$$

avec

$$(34\ bis)\qquad (\sigma_h)^2 = \sum_i \Pi_{hi}(r_i - M)^2 + 2\sum_i \sum_j \Pi_{hi} s_{ij}(r_i - M)(r_j - M)$$

Enfin, dans le cas semi-régulier, qui est compris dans le cas précédent, les σ_h sont égaux entre eux et l'on a :

$$(35)\qquad \lim_{n\to\infty} \text{moy}\ n|\rho_h^{(n)}|^2 = \sigma^2,$$

avec

$$(36)\qquad \sigma^2 = \sum_i \Pi_i(r_i - M)^2 + 2\sum_i \sum_j \Pi_i s_{ij}(r_i - M)(r_j - M)$$

En général, σ n'est pas nul et dans le cas semi-régulier $|\rho_h^{(n)}|^2$ est en moyenne de l'ordre de $\frac{1}{n}$. Mais σ peut être nul et alors $|\rho_h^{(n)}|^2$ est un infiniment petit d'ordre inférieur en moyenne à $\frac{1}{n}$.

On pourrait se demander, s'il serait possible, en perfectionnant le raisonnement, de remplacer la limite en moyenne par la limite au sens ordinaire dans ces dernières formules. Il ne serait pas possible, en tout cas, d'y arriver en faisant ce remplacement pour $S_{ij}^{(n)}$. Autrement dit, il est certain que $S_{ij}^{(n)}$ ne converge pas toujours au sens ordinaire. Le lecteur pourra s'en assurer, en effet, par des calculs simples dans le cas où, supposant $r = 2$, on prend

$$p_{ij} = \begin{cases} 0 & \text{si}\ i = j, \\ 1 & \text{si}\ i \neq j, \end{cases}$$

cas où $S_{ij}^{(n)}$ converge quand n croît par valeurs de même parité vers une limite qui dépend de cette parité. Et pourtant, c'est déjà là un cas semi-régulier, où les Π_{ij} sont tous égaux à $\frac{1}{2}$. Il faut toutefois observer que dans ce cas $n[\delta_h^{(n)}]^2$ a, malgré tout, une limite unique.

Cette observation invite à chercher à résoudre la question sans le concours des $S_{ij}^{(n)}$. M. Dœblin nous a informé qu'on peut y arriver en utilisant un résultat établi, au moyen de l'inégalité de Schwarz, plus loin pages 151, 152. Il en résultera que dans le cas semi-régulier, et même dans le cas d'une variable aléatoire semi-régulière, on peut supprimer les mots « en moyenne » dans les formules respectives (34) et (35).

Valeurs moyennes des fréquences. — D'après la formule (42) de la page 72, on voit que $\mathfrak{M}F_{hk}^{(n,N)}$ converge toujours au moins au sens de Cesaro vers Π_{hk}. Et d'après la formule (43) de la page 73, *la valeur moyenne* $\mathfrak{M}f_{hk}^{(n)}$ *de la fréquence avec laquelle se présente l'état* E_k *dans le cours de n épreuves à partir de l'état* E_k *converge toujours vers une limite* Π_{hk} (qu'on soit ou non dans le cas régulier). Et dans le cas semi-régulier [celui où l'unité est racine simple de $\Delta(s)$], cette limite est une quantité Π_k indépendante de l'état initial E_h.

La quantité $\mathfrak{M}f_{hk}^{n} - \Pi_{hk}$ est toujours infiniment petite avec $\frac{1}{n}$ et même d'ordre au moins égal à celui de $\frac{1}{n}$. Plus précisément le produit $n[\mathfrak{M}f_{hk}^{(n)} - \Pi_{hk}]$ est toujours borné, converge en moyenne vers une limite s_{hk} et même dans le cas où les $P_{hk}^{(n)}$ convergent au sens ordinaire, ce produit converge vers s_{hk} au sens ordinaire.

Dispersion des fréquences. — Il faut, dans les résultats précédents, remplacer, comme à la page 73, $Y_h^{(t)}$ par $U^{(t)}$, $A_h^{(n)}$ par f_{hk}^{n}, M_h par Π_{hk}, x_i par 0 si $i \neq k$ et 1 si $i = k$ [1], $\rho_h^{(n)}$ par $\mathcal{E}_{hk}^{(n)}$ et δ_{hk}^{n} par $\mathcal{G}_{hk}^{n}$. $\lambda_h^{(n)}$ devient l'écart quadratique du nombre $U^{(n)}$ avec Π_{hk} et son carré converge au sens de Cesaro vers ce que devient λ_h^2, soit

$$\Pi_{hk}(1 - \Pi_{hk})$$

Les écarts quadratiques moyens $\mathcal{E}_{hk}^{(n)}$ *et* $\mathcal{G}_{hk}^{(n)}$ de la fréquence f_{hk}^{n} avec sa moyenne $\Pi_{hk}^{(n)}$ et avec la limite Π_{hk} de celle-ci *ont toujours une même limite* w_{hk} (ce que devient ici W_h), et l'on a

$$(w_{hk})^2 = \Pi_{hk}\left(\Pi_{kk} - \sum_i \Pi_{hi}\Pi_{ik}\right),$$

et en vertu de la formule (8) de la page 110, *on a* aussi

$$(w_{hk})^2 = \Pi_{hk}(\Pi_{kk} - \Pi_{hk})$$

On voit que w_{hk} est nul dans le cas régulier et même dans le cas semi-régulier. Dans ce cas tous les w_{hk} sont nuls. On verra (p. 173) que si

---

[1] On voit qu'il suffit de se placer dans un cas oscillant tel que l'exemple de la page 120 pour rencontrer un cas où $\mathfrak{M}A_h^{(n)} = \mathfrak{M}f_{hk}^{(n)}$ converge (vers π_{hk}) tandis que $\mathfrak{M}Y_h^{(n)} = \mathfrak{M}U^{(n)} = P_{hk}^{(n)}$ n'a pas de limite

l'on n'est pas dans le cas semi-régulier, l'un au moins des Π_{hk} est nul, et par conséquent l'un au moins des w_{hk} est encore nul.

Cependant quand on n'est pas dans le cas semi-régulier, les w_{hk} ne sont pas nécessairement tous nuls, comme le montre le cas où le tableau des p_{ik} est

$$\begin{Vmatrix} 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 1 \end{Vmatrix}.$$

On s'assure facilement qu'alors $P_{ik}^{(n)} = p_{ik}$ quel que soit n et que

$$(w_{21})^2 = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4} \neq 0$$

Comme w_{hk} est un cas particulier de W_h, il en résulte aussi que W_h n'est pas toujours nul (p. 137).

La quantité $\{[\mathcal{G}_{hk}^{(n)}]^2 - (w_{hk})^2\}$ est un infiniment petit d'ordre au moins égal à celui de $\frac{1}{n}$. Plus précisément, le produit $n\{[\mathcal{G}_{hk}^{(n)}]^2 - (w_{hk})^2\}$ est toujours borné et même tend en moyenne vers une limite

Cette limite est obtenue en remplaçant dans (33), $(\lambda_h)^2$ par $(1 - \Pi_{hk})\Pi_{hk}$ et x_i par 0 si $i \neq k$, 1 si $i = k$, donc M_h par Π_{hk}, soit

$$(1 - \Pi_{hk})\Pi_{hk} - w_{hk}^2 + E = \Pi_{hk}(1 - \Pi_{kk}) + E,$$

E étant ce que devient

$$2\sum_i (x_i - M_h) \sum_j (\Pi_{ij} s_{hi} + \Pi_{hi} s_{ij})(x_j - M_h)$$

La sommation en j donne

$$\Pi_{ik} s_{hi} + \Pi_{hi} s_{ik} - \Pi_{hk}\left[\left(\sum_j \Pi_{ij}\right) s_{hi} + \Pi_{hi}\left(\sum_j s_{ij}\right)\right].$$

Mais, en vertu des formules de la page 116, le crochet se réduit à s_{hi}. Donc E est ce que devient

$$2\sum_i (x_i - M_h)(\Pi_{ik} s_{hi} + \Pi_{hi} s_{ik} - s_{hi}\Pi_{hk}),$$

soit

$$2(\Pi_{kk}s_{hk}+\Pi_{hk}s_{kk}-s_{hk}\Pi_{hk})-2\Pi_{hk}\left[\sum_i s_{hi}\Pi_{ik}+\sum_i \Pi_{hi}s_{ik}-\Pi_{hk}\sum_i s_{hi}\right]$$

Le nouveau crochet est encore nul en vertu des formules de la page 116, de sorte que E se réduit à la parenthèse. D'où

$$(37)\qquad \lim_{n\to\infty} \text{moy}\,\{n[(\mathcal{G}_{hk}^{(n)})^2-(\varpi_{hk})^2]\}$$
$$=\Pi_{hk}(1-\Pi_{kk})+2(\Pi_{kk}-\Pi_{hk})s_{hk}+2\Pi_{hk}s_{kk}.$$

Dans le cas semi-régulier, ou même, plus généralement, quand $\varpi_{hk}=0$, c'est-à-dire quand

$$\Pi_{hk}=\Pi_{kk} \text{ ou zéro},$$

le carré de l'écart quadratique moyen $\mathcal{E}_{hk}^{(n)}$ de la fréquence $f_{hk}^{(n)}$ est en moyenne infiniment petit avec $\frac{1}{n}$ et même d'un ordre au moins égal à celui de $\frac{1}{n}$.

Non seulement $n[\mathcal{E}_{hk}^{(n)}]^2$ reste borné, mais ce produit tend en moyenne vers une limite qu'on peut appeler $(\nu_{hk})^2$. Celle-ci est égale à l'expression (37), mais se simplifie en y supposant $\varpi_{hk}=0$. Ou bien $\Pi_{hk}=0$, et alors

$$(\nu_{hk})^2=2\Pi_{kk}s_{hk},$$

ou bien (ce qui a lieu en particulier dans le cas semi-régulier) $\Pi_{hk}=\Pi_{kk}$, et alors

$$(\nu_{hk})^2=\Pi_{kk}(1-\Pi_{kk}+2s_{kk}).$$

En particulier, dans le cas semi-régulier, on a pour l'écart quadratique moyen $\mathcal{E}_{hk}^{(n)}$ de la fréquence $f_{hk}^{(n)}$,

$$(37\,bis)\qquad \lim_{n\to\infty} \text{moy.}\,\{n[\mathcal{E}_{hk}^{(n)}]^2\}=\Pi_k\{1-\Pi_k+2s_{kk}\}.$$

En particularisant X(E) comme page 141, on montre que dans le cas semi-régulier on peut supprimer les mots « en moyenne » dans (37) et (37 *bis*).

c — Limite de la loi réduite de répartition des sommes de variables enchaînées.

Rappel d'un cas simple. — Nous allons étudier le comportement asymptotique de la loi de répartition de la moyenne arithmétique

$$A_h^{(n)} = \frac{1}{n}\left[X_h^{(1)} + \ldots + X_h^{(n)}\right].$$

Considérons d'abord le cas simple où chaque état E_i a une probabilité constante p_i, c'est-à-dire où p_{ki} est indépendant de k. Or,

$$A_h^{(n)} = \sum_j f_j^{(n)} x_j,$$

où $f_j^{(n)}$ est la fréquence de E_j en n épreuves et, en posant $M = \Sigma p_j x_j$, on a

$$A_h^{(n)} - M = \sum_j [f_j^{(n)} - p_j] x_j$$

Chacune des variables aléatoires $f_j^{(n)} - p_j$ a une loi réduite de répartition, qui, d'après le théorème de Moivre, converge vers la seconde loi de Laplace, et l'on en conclut facilement, comme il est connu, que $A_k^{(n)} - M$, qui en est une combinaison linéaire, est dans le même cas. L'hypothèse d'une probabilité constante pour E_i, rentrant comme cas particulier dans ce qui va suivre, nous nous contenterons de ces indications.

Cas des variables enchaînées. — Divers auteurs ont cherché à étendre ce théorème au cas où les p_{ki} dépendent des k. Ils y ont réussi sous des conditions de généralité croissante, Markoff [15] en supposant les p_{ki} tous $\neq 0$ et le nombre des valeurs x_i égal à deux; M. S. Bernstein [1] en se plaçant dans le cas positivement régulier; M. Dœblin [2] a réussi à étendre la méthode de S. Bernstein, et en analysant aussi le cas singulier, a indiqué certains cas où la loi réduite limite est plus complexe.

Méthode de Schulz et Mihoc. — En même temps, MM. Schulz [1. 2], Romanovsky [1], puis Schulz [4] et Mihoc [2] appliquaient dans le même but une méthode différente : la méthode classique qui consiste à calculer les limites des moments de la loi réduite et à montrer qu'ils tendent vers ceux de la seconde loi de Laplace. Tout étant ramené au calcul des parties principales des moments, M. Romanovsky [1] y procède directement par des moyens analogues, pour les ordres supérieurs, à ceux qui ont été indiqués, page 76, pour le moment d'ordre deux. Malgré la complication accrue, sa grande puissance de calcul lui permet d'arriver ainsi le premier à établir complètement la loi de Gauss dans le cas positivement régulier. Les deux derniers auteurs, opérant indépendamment l'un de l'autre, effectuent, au contraire, indirectement ce calcul en le réduisant à l'étude du comportement asymptotique des solutions d'un système d'équations aux différences. Nous rappellerons les résultats de ce dernier problème qui est purement algébrique, dans la Note C, page 278, et nous en appliquerons ici les résultats au problème actuel.

Nous introduirons donc dans la méthode de MM. Schulz et Mihoc des précisions et des modifications — qui auront, entre autres, l'avantage d'éliminer l'emploi de déterminants toujours encombrants, de prouver rigoureusement l'existence des développements limités considérés et de faciliter le calcul explicite de certaines limites.

A. — Comportement asymptotique des moments dans le cas général

Relation de récurrence entre moments. — Soit $Y_h^{(n)}$ la valeur prise après n épreuves par la variable aléatoire $X(E)$ quand le système était initialement à l'état E_h : posons

$$S_j^{(n)} = Y_j^{(1)} + \ldots + Y_j^{(n)} = n A_j^{(n)}, \qquad M_j^{(n)} = \mathfrak{M} A_j^{(n)}.$$

On va établir des relations de récurrence entre les moments

$$\mathfrak{M}_{jn}^{(\nu)} = \mathfrak{M}[Y_j^{(1)} + \ldots + Y_j^{(n)}]^\nu,$$

qu'on peut désigner par

$$\mathfrak{M}_{E_j}[X^{(1)} + \ldots + X^{(n)}]^\nu.$$

Si aux épreuves successives se réalisent les états E_i, E_{i_1}, ...,

$E_{i_n}, \ldots$, l'espérance mathématique correspondante de $|X_j^{(1)} + \ldots + X_j^{(n)}|^\nu$ est

$$p_{ji} p_{ii_1} \ldots p_{i_{n-2} i_{n-1}} (x_i + x_{i_1} + \ldots + x_{i_{n-1}})^\nu.$$

On a donc

$$(38) \quad \mathfrak{M}_{jn}^{(\nu)} = \sum_i p_{ji} \sum_{i_1 i_2 \ldots i_{n-1}} p_{ii_1} \ldots p_{i_{n-2} i_{n-1}} (x_{i_1} + \ldots + x_{i_{n-1}})^\nu + \ldots$$
$$+ C_\nu^\gamma \sum_i p_{ji} x_i^\gamma \left\{ \sum_{i_1 i_2 \ldots i_{n-1}} p_{ii_1} \ldots p_{i_{n-2} i_{n-1}} (x_{i_1} + \ldots + x_{i_{n-1}})^{\nu-\gamma} \right\} + \ldots$$
$$+ \sum_i p_{ji} x_i^\nu \sum_{i_1 i_2 \ldots i_{n-1}} p_{ii_1} \ldots p_{i_{n-2} i_{n-1}}.$$

D'où, enfin,

$$(38\ bis) \quad \mathfrak{M}_{jn}^{(\nu)} = \sum_i p_{ji} \mathfrak{M}_{i,n-1}^{(\nu)} + \ldots + C_\nu^\gamma \sum_i p_{ji} x_i^\gamma \mathfrak{M}_{i,n-1}^{(\nu-\gamma)} + \ldots + \sum_i p_{ji} x_i^\nu.$$

On a vu que $M_j^{(n)} = \frac{1}{n} \mathfrak{M}_{jn}^{(1)}$ converge toujours vers

$$M_j = \sum_i \Pi_{ji} x_i.$$

Au lieu des $\mathfrak{M}_{jn}^{(\nu)}$, il y aura donc intérêt pour les applications à considérer les

$$\mathfrak{N}_{jn}^{(\nu)} = \mathfrak{M}_{E_j} \{ (X^{(1)} - M_h) + \ldots + (X^{(n)} - M_h) \}^\nu = n^\nu \mathfrak{M} | A_j^{(n)} - M_h |^\nu,$$

et plus particulièrement les $\mathfrak{N}_{h,n}^{(\nu)}$. On obtiendra la relation de récurrence entre les $\mathfrak{N}_{j,n}^{(\nu)}$ en remplaçant dans (38) les $\mathfrak{M}$ par des $\mathfrak{N}$ et les x_i par des $z_i = x_i - M_h$, soit

$$(39) \quad \mathfrak{N}_{kn}^{(\nu)} - \sum_i p_{ki} \mathfrak{N}_{i,n-1}^{(\nu)} = f_k^{(\nu)}(n-1),$$

avec

$$(39\ bis) \quad f_k^{(\nu)}(n) = C_\nu^1 \sum_i p_{ki} z_i \mathfrak{N}_{i,n}^{(\nu-1)} + \ldots + C_\nu^\gamma \sum_i p_{ki} z_i^\gamma \mathfrak{N}_{i,n}^{(\nu-\gamma)} + \ldots + \sum_i p_{ki} z_i^\nu.$$

Parties principales des moments. — On a

$$\mathfrak{N}_{kn}^{(1)} = n \mathfrak{M} [A_k^{(n)} - M_h] = n [M_k^{(n)} - M_k] + n (M_k - M_h)$$

et, par suite, d'après (24 *bis*), page 133,

$$\mathfrak{N}^{(1)}_{k,n} = n(\mathrm{M}_k - \mathrm{M}_h) + \sum_l s_{kl} z_l + \varphi^{(n)}_k, \tag{40}$$

où les $\varphi^{(n)}_k$ sont solutions du système homogène

$$u_k(n) - \sum_l p_{kl} u_l(n-1) = 0, \tag{H}$$

et convergent en moyenne vers zéro

On a donc

$$\varphi^{(n)}_k = \sum_j \mathrm{P}^{(n-1)}_{kj} \varphi^{(1)}_j,$$

et en passant à la limite en moyenne

$$0 = \sum_j \Pi_{kj} \varphi^{(1)}_j,$$

d'où

$$\varphi^{(n)}_i = \sum_j [\mathrm{P}^{(n-1)}_{ij} - \Pi_{ij}] \varphi^{1}_j \tag{40 bis}$$

D'après la formule (13) de la page 113, on retrouve ainsi la formule (24 *ter*).

Généralisons (40). On a d'abord

$$\mathfrak{N}^{(1)}_{k,n} = n(\mathrm{M}_k - \mathrm{M}_h) + \mathrm{A}^{(1)}_k(n), \tag{40 ter}$$

où $\mathrm{A}^{(1)}_k(n)$ est une fonction asymptotiquement périodique de n, de période N. Supposons qu'on ait pour $t < \nu$,

$$\mathfrak{N}^{(t)}_{k,n} = n^t \mathrm{L}_{kt} + n^{t-1} \mathrm{A}^{(t)}_k(n), \tag{41}$$

où $\mathrm{A}^{(t)}_k(n)$ est une fonction asymptotiquement périodique de n, de période N et où L_{kt} est une constante. Alors, comme

$$f^{(\nu)}_k(n) = \mathrm{C}^1_\nu \sum_l p_{kl} z_l \mathfrak{N}^{(\nu-1)}_{l,n} + \mathrm{C}^2_\nu \sum_l p_{kl} z^2_l \mathfrak{N}^{(\nu-2)}_{l,n} + \dots,$$

on peut écrire

$$f^{(\nu)}_k(n) = n^{\nu-1} \mathrm{L}'_{k\nu} + n^{\nu-2} \mathrm{D}^{(\nu)}_k(n),$$

où

$$\mathrm{L}'_{k\nu} = \nu \sum_l p_{kl} z_l \mathrm{L}_{l,\nu-1},$$

et où

$$D_k^{(\nu)}(n) = C_\nu^1 \sum_l p_{kl} z_l A_l^{(\nu-1)}(n) + C_\nu^2 \sum_l p_{kl} z_l^2 L_{l,\nu-2}$$
$$+ \frac{1}{n} C_\nu^2 \sum_l p_{kl} z_l^2 A_l^{(\nu-2)}(n)$$
$$+ \frac{1}{n} C_\nu^3 \sum_l p_{kl} z_l^3 L_{l,\nu-3} + \ldots + \frac{1}{n^{\nu-2}} \sum_l p_{kl} (z_l)^\nu,$$

et par suite $D_k^{(\nu)}(n)$ est une fonction asymptotiquement périodique de période N.

Il en résulte que $\mathfrak{N}_{kn}^{(\nu)}$ est la somme de deux termes qui sont les solutions du système (H) où l'on placerait successivement au second membre les deux termes de $f_k^{(\nu)}(n)$.

Alors, on peut en conclure (note C, p. 279, 280) que la formule (41) est vraie pour $t = \nu$ et que

$$L_{k\nu} = \frac{1}{\nu} \sum_l \Pi_{kl} L'_{l\nu}.$$

Des lors, l'hypothèse faite pour $t < \nu$ étant réalisée pour $\nu = 2$, le sera pour toute valeur de ν. Finalement : *pour toute valeur de ν, on a*

$$\mathfrak{N}_{k,n}^{(\nu)} = n^\nu L_{k\nu} + n^{\nu-1} A_k^{(\nu)}(n), \tag{41 bis}$$

où $A_k^{(\nu)}(n)$ est une fonction asymptotiquement périodique de n et où $L_{k\nu}$ est une constante vérifiant la relation de récurrence

$$L_{k\nu} = \sum_j \Pi_{kj} z_j L_{j,\nu-1}, \tag{42}$$

avec

$$L_{k1} = M_k - M_h. \tag{43}$$

En particulier, pour $\nu = 2$:

$$\mathfrak{N}_{k,n}^{(2)} = n^2 L_{k2} + n A_k^{(2)}(n),$$

de sorte que $\frac{\mathfrak{N}_{kn}^{(2)}}{n^2}$ converge au sens ordinaire vers

$$L_{k2} = \sum_j \Pi_{kj} z_j (M_j - M_h) = \sum_j \sum_l \Pi_{kj} \Pi_{jl} (x_j - M_h) (x_l - M_h).$$

Ainsi

$$\lim_{n\to\infty}[\delta_h^{(n)}]^2 = \lim_{n\to\infty}\left\{\frac{1}{n}\mathfrak{M}_{\mathrm{L}_h}[(\mathrm{X}^{(1)}-\mathrm{M}_h)+\ldots+(\mathrm{X}^{(n)}-\mathrm{M}_h)]^2\right\}$$
$$=\sum_{l}\sum_{h}\Pi_{hj}\Pi_{jl}(x_j-\mathrm{M}_h)(x_l-\mathrm{M}_l)=[\mathrm{W}_h]^2.$$

c'est le résultat de la page 138 valable, ici encore, dans le cas singulier le plus général comme dans le cas régulier.

Plus généralement, on tire de (42) et de (43)

$$(44)\qquad \mathrm{L}_{k\nu}=\sum_{i_\nu,\ldots,i_1}\Pi_{ki_\nu}\Pi_{i_\nu i_{\nu-1}}\cdots\Pi_{i_2 i_1}z_{i_\nu}\cdots z_{i_2}z_{i_1}$$
$$=\sum_{i_\nu}\Pi_{ki_\nu}z_{i_\nu}\left[\sum_{i_{\nu-1}}\cdots\left[\sum_{i_1}\Pi_{i_2 i_1}z_{i_1}\right]\cdots\right].$$

Cas d'une variable aléatoire semi-régulière — Considérons maintenant le cas semi-régulier ou tout d'abord, plus généralement, le cas où les x_i sont tels que les $\mathrm{M}_k=\Sigma\Pi_{ki}x_i$ soient indépendants de k. La dernière expression de $\mathrm{L}_{k\nu}$, ou

$$\sum_{i_1}\Pi_{i_2 i_1}z_{i_1}=\mathrm{M}_{i_2}-\mathrm{M}_k$$

montre qu'alors tous les $\mathrm{L}_{k\nu}$ sont nuls. $\mathfrak{M}_{k,n}^{(\nu)}$ est comme $\mathrm{A}_k^{(\nu)}(n)$ une fonction asymptotiquement périodique, de période N.

Ainsi, *dans le cas où les quantités*

$$\mathrm{M}_k=\sum_j\Pi_{kj}x_j$$

sont égales à une même valeur M — et en particulier dans le cas semi-régulier, — *les moments d'ordre* ν *sont au plus de l'ordre de* $n^{\nu-1}$. Plus précisément, ils sont *de la forme*

$$\mathfrak{M}_{k,n}^{(\nu)}=n^{\nu-1}\mathrm{A}_k^{(\nu)}(n),$$

où $\mathrm{A}_k^{(\nu)}(n)$ *est une fonction asymptotiquement périodique de période* N.

En particulier, posant

$$\lim_{n\to\infty}\mathrm{moy.}\,\mathrm{A}_k^{(\nu)}(n)=\mathrm{K}_{k\nu},$$

on aura comme plus haut la relation de récurrence

$$K_{k\nu} = \frac{\nu}{\nu - 1} \sum_j \Pi_{kj} z_j K_{j,\nu-1},$$

car si l'on écrit

$$f_k^{(\nu)}(n) = n^{\nu-2} D_k^{(\nu)}(n),$$

on a, d'après la Note C,

$$\lim_{n \to \infty} \text{moy. } A_k^{(\nu)}(n) = \frac{1}{\nu - 1} \sum_i \Pi_{ki} \Big[\lim_{n \to \infty} \text{moy. } D_i^{(\nu)}(n) \Big]$$

avec

$$D_k^{(\nu)}(n) = \nu \sum_i p_{ki} z_i A_i^{(\nu-1)}(n) + \ldots + \frac{1}{n^{\nu-2}} \sum_i p_{ki} (z_i)^\nu$$

Cependant, ceci suppose $\nu > 2$; on voit, en effet, que la limite en moyenne de $D_k^{(\nu)}(n)$, qui est en général égal à $\nu \sum_i p_{ki} z_i K_{i,\nu-1}$, est pour $\nu = 2$, la somme de $2 \sum_i p_{ki} z_i K_{i1}$ et de $\sum_i p_{ki} (z_i)^2$. De sorte que

$$K_{k2} = 2 \sum_j \Pi_{kj} z_j K_{j1} + \sum_j \Pi_{kj} z_j^2,$$

et comme $K_{j1} = \sum_i s_{ji} z_i$, on voit qu'on a non seulement

$$\mathfrak{N}_{kn}^{(2)} = n A_k^{(2)}(n), \tag{45}$$

mais

$$\lim_{n \to \infty} \text{moy. } \frac{\mathfrak{N}_{kn}^{(2)}}{n} = \sum_j \Pi_{kj} z_j^2 + 2 \sum_j \sum_i \Pi_{kj} s_{ji} z_i z_j = (\sigma_k)^2 \tag{45 bis}$$

Dans le cas semi-régulier, on voit que $(\sigma_k)^2$ devient indépendant de k et égal à

$$\sigma^2 = \sum_i \Pi_j z_j^2 + 2 \sum_i \sum_j \Pi_j s_{ji} z_i z_j.$$

Dans ce cas

$$\lim_{n \to \infty} \text{moy. } \frac{\mathfrak{N}_{kn}^{(3)}}{n^2} = K_{k3} = \frac{3}{2} \sum_j \Pi_{kj} z_j K_{j2} = \frac{3}{2} \Big[\sum_j \Pi_j z_j \Big] \sigma^2 = 0. \tag{46}$$

On peut encore préciser ces résultats. En effet, revenons au cas d'une variable aléatoire semi-régulière et tenons compte des for-

mules (24 *ter*) et (40) : $\mathfrak{M}^{(1)}_{kn}$ ne diffère d'une fonction de n de période N que par le terme général d'une série absolument convergente. Il en sera alors nécessairement de même de $f^{(2)}_k(n)$. Or, on sait (Note C, p. 280) que, dans ce cas, toute solution de (39), pour $\nu = 2$, est de la forme

$$\mathfrak{M}^{(2)}_{kn} = n\varpi_k(n) + \mathrm{H}_k(n), \tag{47}$$

où $\varpi_k(n)$ est de période N et $\mathrm{H}_k(n)$ est borné. [La précision apportée à (45) consiste en ce que, $\mathrm{A}^{(2)}_k(n) = \varpi_k(n) + \dfrac{\mathrm{H}_k(n)}{n}$ différant d'une fonction périodique d'un infiniment petit, on sait maintenant celui-ci d'ordre au moins égal à celui de $\frac{1}{n}$.]

Un raisonnement qui nous a été communiqué par M. Dœblin, permet d'aller plus loin. On a, d'une part,

$$\mathfrak{M}^{(2)}_{k,n+1} - \mathfrak{M}^{(2)}_{k,n} = n[\varpi_k(n+1) - \varpi_k(n)] + (\text{quantité bornée})$$

et, d'autre part,

$$\mathfrak{M}^{(2)}_{k,n+1} - \mathfrak{M}^{(2)}_{k,n} = \mathfrak{M}[\mathrm{S}^{(n+1)}_k - (n+1)\mathrm{M}_h]^2 - \mathfrak{M}[\mathrm{S}^{(n)}_k - n\mathrm{M}_h]^2 \tag{48}$$

avec

$$\mathrm{S}^{(n+1)}_k = \mathrm{S}^{(n)}_k + \mathrm{Y}^{(n+1)}_k.$$

Le second membre de (48) est donc

$$\mathfrak{M}[\mathrm{Y}^{(n+1)}_k - \mathrm{M}_h]^2 + 2\{\mathfrak{M}[\mathrm{S}^{(n)}_k - n\mathrm{M}_h]\}\{\mathfrak{M}\mathrm{Y}^{(n+1)}_k - \mathrm{M}_h\}$$

Or, $\mathrm{Y}^{(n+1)}_k$ ne pouvant prendre que les r valeurs x_i, la première moyenne a, quand n, k et h varient, une borne supérieure $\mathrm{G}^2 \geq 0$, et en vertu de l'inégalité de Schwarz ([1]), le second terme est, en valeur absolue, au plus égal à

$$2\sqrt{\mathfrak{M}^{(2)}_{kn}}\,\mathrm{G}$$

On a donc

$$\left|\varpi_k(n+1) - \varpi_k(n) + \frac{\text{quantité bornée}}{n}\right| \leq \frac{\mathrm{G}^2}{n} + 2\frac{\mathrm{G}\sqrt{n\varpi_k(n) + \mathrm{H}_k(n)}}{n}.$$

En remplaçant n par $n + t\mathrm{N}$ et faisant croître l'entier t, on obtient

---

([1]) *Voir*, par exemple, Fréchet [188, p. 70]

à la limite

$$\varpi_k(n+1) - \varpi_k(n) = 0,$$

c'est-à-dire que $\varpi_k(n)$ est même une constante. En vertu de (47), cette constante est la limite de $\frac{1}{n}\mathfrak{M}_{kn}^{(2)}$. Or, on vient de voir page 150, que ce rapport converge, au moins en moyenne, vers $K_{k2} = (\sigma_k)^2$. Il est donc maintenant établi que ce rapport converge aussi au sens ordinaire vers $(\sigma_k)^2$.

Ainsi, comme nous l'avions annoncé, on peut supprimer le mot « moy. » dans les formules (34) et (35).

De sorte que *dans le cas d'une variable aléatoire semi-régulière, on a*

$$(34\ ter)\qquad \lim n[\delta_k^{(n)}]^2 = \lim_{n\to\infty} \frac{\mathfrak{M}_{kn}^{(2)}}{n} = (\sigma_k)^2$$

$$= \sum_j \Pi_{kj}(x_j - M)^2 + 2\sum_j\sum_i \Pi_{kj} s_{ji}(x_j - M)(x_i - M)$$

On a, en particulier, *dans le cas semi-régulier*,

$$(35\ ter)\qquad \lim n[\delta_k^{(n)}]^2 = \lim_{n\to\infty} \frac{\mathfrak{M}_{kn}^{(2)}}{n} = \sigma^2$$

$$= \sum_j \Pi_j(x_j - M)^2 + 2\sum_j\sum_i \Pi_j s_{ji}(x_j - M)(x_i - M)$$

Ces formules fournissent un premier renseignement sur les comportements de $A_k^{(n)}$ et $f_{hk}^{(n)}$. Puisque $\mathfrak{M}[A_k^{(n)} - M]^2 = \frac{\mathfrak{M}_{kn}^{(2)}}{n^2}$, la formule (34 *ter*) montre que $\mathfrak{M}[A_k^{(n)} - M]^2$ tend vers zéro. Il en résulte que dans le cas d'une variable semi-régulière, $A_k^{(n)}$ converge vers M « en moyenne quadratique » (*voir* cet Ouvrage, Premier Livre, p. 205) et par suite aussi « en probabilité » (*id.*, p. 164). En particularisant X(E) comme à la page 141, on voit que : dans le cas semi-régulier, la fréquence (aléatoire) $f_{hk}^{(n)}$ converge vers Π_k en moyenne quadratique et par suite en probabilité.

Ces résultats, conséquences immédiates de (34 *ter*) seront améliorés plus loin, page 159.

En revenant au calcul des moments dans le cas d'une variable aléatoire semi-régulière, comme on a

$$(47\ bis)\qquad \mathfrak{M}_{kn}^{(2)} = n\sigma_k^2 + (\text{quantité bornée}),$$

on voit que . *dans le cas d'une variable aléatoire semi-régulière, pour que la valeur moyenne de* $\{(Y_k^{(1)} - M) + \ldots + (Y_k^{n} - M)\}^2$ *soit bornée quand n croît, il faut et il suffit que* $\sigma_k = 0$.

D'autre part, on a, en vertu de (47 *bis*).

$$f_k^{(3)}(n) = 3\sum_l p_{kl}\, z_l\, n\, \sigma_l^2 + (\text{quantité bornée})$$

En remplaçant successivement $f_{kin}^{(3)}$ dans (39) par ses deux termes, on voit, d'après la Note C, que $\mathfrak{M}_{kn}^{(3)}$ est la somme d'un polynome $K_{k3} n^2 + \ldots$, d'une quantité bornée et d'une quantité égale a n fois une quantité bornée, d'où

$$\mathfrak{M}_{kn}^{(3)} = \frac{n^2}{2}\sum_l \Pi_{kl}\, 3\sum_i p_{li}\, z_i \sigma_i^2 + n(\text{quantité bornée})$$

Donc

$$\mathfrak{M}_{kn}^{(3)} = \frac{3n^2}{2}\Sigma \Pi_{kl}\, z_l \sigma_l^2 + n(\text{quantité bornée})$$

Ainsi $\dfrac{\mathfrak{M}_{kn}^{(3)}}{n^2}$ a une limite dans le cas d'une variable aléatoire semi-régulière. On a vu plus haut que, dans le cas semi-régulier, cette limite est nulle. Mais d'après M. Dœblin [3], en considérant séparément les termes de $\sum \Pi_{kl} z_l \sigma_l^2$ relatifs à chacun des groupements indécomposables définis plus loin (p. 173, 174), on peut montrer que cette limite est encore nulle dans le cas plus général d'une variable aléatoire semi-régulière, de sorte qu'on peut aussi supprimer le mot « moy » dans (46), et même, on a ce résultat plus précis que $\dfrac{\mathfrak{M}_{kn}^{(3)}}{n}$ est borné.

On trouvera dans un mémoire de M. Mihoc [2] et, ultérieurement, mais plus complètement dans la Thèse de M. Dœblin [4], une étude plus étendue du comportement asymptotique des moments des divers ordres dans le cas singulier. Nous allons nous borner dans ce qui suit à l'étude du cas régulier.

B. — Comportement asymptotique des moments dans le cas régulier.

Développements limités des moments — Nous pourrions aller plus vite en utilisant les calculs précédents. Mais comme le cas régulier est le plus important et le plus simple, nous procéderons directement.

Revenons à la formule (40 *bis*).

Il en résulte que, dans le cas régulier, la série $\sum_n \varphi_i^{(n)}$ converge absolument et, de même que $(P_{ij}^{(n-1)} - P_j)$, $\varphi_i^{(n)}$ tend vers zéro exponentiellement.

Nous allons généraliser ce résultat. Nous voyons que, au moins pour $\nu = 2$, on peut écrire pour $t < \nu$, dans le cas régulier, $\mathfrak{M}_{k,n}^{(t)}$ sous la forme

$$\mathfrak{M}_{k,n}^{(t)} = \varphi_{k,t}^{(n)} + A_{k,t} + B_{k,t}n + \quad + K_{k,t}n^{t-1}, \tag{49}$$

où les A, ..., K sont des constantes par rapport à n et où $\varphi_{k,t}^{(n)}$ tend vers zéro exponentiellement.

On a alors, d'après (39 *bis*),

$$f_k^{(\nu)}(n) = C_\nu^1 \sum_l p_{kl} z_l [\varphi_{l,\nu-1}^{(n)} + A_{l,\nu-1} + .. + K_{l,\nu-1}n^{\nu-2}] + \quad + \sum_l p_{kl} z_l^\nu$$
$$= \eta_{k,\nu-1}^{(n)} + A'_{k,\nu-1} + \ldots + H'_{k,\nu-1}n^{\nu-2},$$

où les A′, B′, ..., H′ sont des constantes et où $\eta_{k,\nu-1}^{(n)}$ converge vers zéro exponentiellement, ce dernier mot étant essentiel pour la déduction qui va suivre.

Comme les $\mathfrak{M}_{k,n}^{(\nu)}$ vérifient l'équation

$$y_k(n) - \sum_l p_{kl} y_l(n-1) = \eta_{k,\nu-1}^{(n-1)} + A'_{k,\nu-1} + \;..+ H'_{k,\nu-1}(n-1)^{\nu-2},$$

on en conclut (Note C, p. 279) que *les $\mathfrak{M}_{k,n}^{(\nu)}$ admettent un développement limité de la forme* (49), *soit*

$$\mathfrak{M}_{k,n}^{(\nu)} = \varphi_{k,\nu}^{(n)} + A_{k,\nu} + \quad + K_{k,\nu}n^{\nu-1},$$

où les A, ..., K sont des constantes et où $\varphi_{k,\nu}^{(n)}$ converge vers zéro. Et même, $\varphi_{k,\nu}^{(n)}$ converge aussi exponentiellement. D'après ce qui précède, ceci a lieu pour toute valeur entière de ν.

Nous allons maintenant prouver que la partie entière de ce développement est, pour $\nu > 2$, d'un degré inférieur à $\nu - 1$, et qu'on peut préciser.

Moments d'ordre deux. — Dans tous les calculs qui vont suivre, on aura constamment à faire des simplifications basées sur les relations (8 *bis*, 8 *ter*) et (9) des pages 46 et 47 que vérifient les P_l et les s_{ik}.

Formons d'abord le développement limité de $\mathfrak{M}_{k,n}^{(2)}$.

On a

$$\mathfrak{M}_{k,n}^{(2)} = \varphi_{k,2}^{(n)} + A_{k,2} + n B_{k,2}, \tag{50}$$

et, d'après la Note (C), page 279,

$$B_{k,2} = \lim_{n\to\infty} \frac{\mathfrak{M}_{k,n}^{(2)}}{n} = \sum_l P_l F_l,$$

en posant

$$\begin{aligned} F_l &= \lim_{n\to\infty} f_l^{(2)}(n) \\ &= \lim_{n\to\infty}\left[2\sum_i p_{li} z_i \mathfrak{M}_{i,n}^{(1)} + \sum_i p_{li} z_i^2\right] = 2\sum_i p_{li} z_i \sum_j s_{ij} z_j + \sum_i p_{li} z_i^2. \end{aligned}$$

Donc $B_{k,2}$ est une quantité indépendante de k. Calculons-la.

$$B_{k,2} = \sum_l P_l F_l = \sum_i \left\{\sum_l P_l p_{li}\right\}\left\{2 z_i \sum_j s_{ij} z_j + z_i^2\right\},$$

soit

$$\lim_{n\to\infty} \frac{\mathfrak{M}_{k,n}^{(2)}}{n} = 2\sum_i \sum_j P_i s_{ij} z_i z_j + \sum_i P_i z_i^2 = \sigma^2.$$

C'est le résultat obtenu page 79.

En outre, on a $\mathfrak{M}_{kn}^{(2)} = n\sigma^2 + A_{k2} + \varphi_{k2}$, de sorte que dans le cas où $\sigma = 0$, $\mathfrak{M}_{kn}^{(2)}$ converge vers une limite finie A_{k2} [dont la valeur est précisée plus loin (p. 163)].

Moments d'ordre trois. — Passons aux moments d'ordre 3 ; on a

$$\mathfrak{M}^{(3)}_{k,n} = \varphi^{(n)}_{k,3} + A_{k,3} + n\,B_{k,3} + n^2 C_{k,3},$$

avec

$$(51) \qquad \begin{aligned} f_k^{(3)}(n) &= 3n\sum_i p_{ki}\,z_i\,\sigma^2 + 3\sum_i p_{ki}\,z_i\,A_{i,2} \\ &\quad + 3\sum_i p_{ki}\,z_i^2 \sum_l s_{il}\,z_l + \sum_i p_{ki}\,z_i^3 + \eta_k(n) \\ &= \eta_k(n) + A'_k + n\,B'_k, \end{aligned}$$

où $\eta_k(n)$ converge exponentiellement vers zéro. Alors on sait (Note C, p. 279) que

$$C_{k,3} = \frac{1}{3}\sum_k P_k B'_k = \frac{1}{3}\,3\sum_i \Big(\sum_k P_k\,p_{ki}\Big) z_i\,\sigma^2 = \frac{3}{3}\Big(\sum_i P_i\,z_i\Big)\sigma^2 = 0$$

Ainsi,

$$\lim \frac{\mathfrak{M}^{(3)}_{k,n}}{n^2} = 0$$

et, par suite, d'après (51), $\dfrac{\mathfrak{M}^{(3)}_{k,n}}{n}$ converge vers une limite, à savoir $B_{k,3}$.

Moments d'ordre supérieur. — Il est maintenant acquis que les expressions

$$\mathfrak{M}^{(1)}_{k,n}, \quad \frac{1}{n}\mathfrak{M}^{(2)}_{k,n}, \quad \frac{1}{n}\mathfrak{M}^{(3)}_{k,n}$$

convergent vers des limites respectives $\theta_k^{(1)}$, $\theta_k^{(2)}$, $\theta_k^{(3)}$ quand n croît. Et même,

$$\theta_k^{(2)} = \sigma^2$$

est indépendant de k.

D'après ce qui précède, on voit qu'il existe au moins une valeur 2, de ν telle que

$$\frac{1}{n^l}\mathfrak{M}^{(2l)}_{k,n} \quad \text{et} \quad \frac{1}{n^l}\mathfrak{M}^{(2l+1)}_{k,n}$$

convergent vers des limites — qu'on peut appeler $\theta^{(2l)}$, $\theta_k^{(2l+1)}$ — pour $l < \nu$. On aura

$$\frac{1}{n^{\nu-1}} f_k^{(2\nu)}(n) = C^1_{2\nu}\sum_i p_{ki}\,z_i\,\frac{\mathfrak{M}^{(2\nu-1)}_{i,n}}{n^{\nu-1}} + C^2_{2\nu}\sum_i p_{ki}\,z_i^2\,\frac{\mathfrak{M}^{(2\nu-2)}_{i,n}}{n^{\nu-1}} + \ldots,$$

d'où

$$\lim_{n\to\infty} \frac{1}{n^{\nu-1}} f_k^{(2\nu)}(n) = F_k^{(2\nu)} = C_{2\nu}^1 \sum_l p_{kl}\, z_l\, \theta_l^{(2\nu-1)} + C_{2\nu}^2 \sum_l p_{kl}\, z_l^2\, \theta^{(2\nu-2)}$$

Donc, d'apres la Note C, $\frac{\mathfrak{N}_{i,n}^{(2\nu)}}{n^\nu}$ converge Et même sa limite étant $\frac{1}{\nu}\sum_k P_k F_k^{(2\nu)}$, est une quantité $\theta^{(2\nu)}$ qui est indépendante de i.

Alors

$$\begin{aligned} f_k^{(2\nu+1)}(n) = {} & C_{2\nu+1}^1 \sum_l p_{kl}\, z_l \{ \theta^{(2\nu)} n^\nu + J_{l,2\nu} n^{\nu-1} + \ldots \} \\ & + C_{2\nu+1}^2 \sum_l p_{kl}\, z_l^2 \{ \theta_l^{(2\nu-1)} n^{\nu-1} + \ldots \} \\ & + C_{2\nu+1}^3 \sum_l p_{kl}\, z_l^3 \{ \theta^{(2\nu-2)} n^{\nu-1} + \ldots \} \\ & + \ldots \\ = {} & F_k n^\nu + G_k n^{\nu-1} + \ldots \end{aligned}$$

Dès lors, dans la formule

$$\mathfrak{N}_{k\,n}^{(2\nu+1)} = J_{k,2\nu+1}\, n^{\nu+1} + I_{k,2\nu+1}\, n^\nu + \ldots$$

on aura d'abord, d'après la Note C,

$$\begin{aligned} J_{k,2\nu+1} &= \frac{1}{\nu+1} \sum_l P_l F_l \\ &= \frac{C_{2\nu+1}^1}{\nu+1} \sum_i \left(\sum_l P_l\, p_{li} \right) z_i\, \theta^{(2\nu)} = \frac{C_{2\nu+1}^1}{\nu+1} \left(\sum_i P_i\, z_i \right) \theta^{(2\nu)} = 0 \end{aligned}$$

Donc, $J_{k,2\nu+1} = 0$, et, par suite, $\frac{\mathfrak{N}_{k,n}^{(2\nu+1)}}{n^\nu}$, convergent vers une certaine limite $\theta_k^{(2\nu+1)} = I_{k,2\nu+1}$.

En résumé, l'hypothese admise pour $t < \nu$ s'étend nécessairement à $t = \nu$ et étant vraie pour $\nu = 2$ est générale. On a même démontré que tous les $\theta_k^{(2\nu)}$ sont indépendants de k Il n'en est pas de même pour les $\theta_k^{(2\nu+1)}$; déjà $\theta_k^{(1)} = \sum_j s_{kj}\, z_j$ dépend en général de k.

On va maintenant établir des relations entre les $\theta^{(2\nu)}$. On a observé

que

$$\theta^{(2\nu)} = \frac{1}{\nu} \sum_k P_k F_k^{(2\nu)},$$

c'est-à-dire

$$(52) \qquad \theta^{(2\nu)} = \frac{1}{\nu} \sum_l \sum_k P_k p_{kl} \{ C_{2\nu}^1 z_l \theta_l^{(2\nu-1)} + C_{2\nu}^2 z_l^2 \theta^{(2\nu-2)} \}$$

$$= 2 \sum_l P_l z_l \theta_l^{(2\nu-1)} + (2\nu - 1) \left[\sum_l P_l z_l^2 \right] \theta^{(2\nu-2)},$$

D'autre part, on a

$$\mathfrak{M}_{k,n}^{(2\nu+1)} - \sum_l p_{kl} \mathfrak{M}_{l,n-1}^{(2\nu+1)} = C_{2\nu+1}^1 \sum_l p_{kl} z_l \mathfrak{M}_{l,n-1}^{(2\nu)} + \ldots$$

Donc, en divisant par n^ν, et passant à la limite

$$\theta_k^{(2\nu+1)} - \sum_l p_{kl} \theta_l^{(2\nu+1)} = (2\nu + 1) \sum_l p_{kl} z_l \theta^{(2\nu)}$$

On voit que les $\theta_k^{(2\nu+1)}$ sont solutions du système

$$y_k - \sum_l p_{kl} y_l = \lambda_k$$

avec

$$\lambda_k = (2\nu + 1) \sum_l p_{kl} z_l \theta^{(2\nu)}.$$

On en conclut (Note C, p. 279) que

$$y_k = A + \lambda_k + \sum_l s_{kl} \lambda_l = A + B_k,$$

où A est indépendant de k et

$$B_k = (2\nu + 1) \sum_l \left\{ p_{kl} + \left[\sum_l s_{kl} p_{ll} \right] \right\} z_l \theta^{(2\nu)}$$

$$= (2\nu + 1) \sum_l (s_{kl} + P_l) z_l \theta^{(2\nu)} = (2\nu + 1) \theta_k^{(1)} \theta^{(2\nu)}$$

En portant l'expression obtenue

$$\theta_k^{(2\nu+1)} = A + (2\nu + 1) \theta_k^{(1)} \theta^{(2\nu)}$$

dans (52), on obtient

$$\theta^{(2\nu+2)} = 2\left(\sum_i P_i z_i\right) A + 2(2\nu+1)\sum_i P_i z_i \theta_i^{(1)} \theta^{(2\nu)} + (2\nu+1)\sum_i P_i z_i^2 \theta^{(2\nu)}$$

$$= (2\nu+1)\left\{2\sum_i P_i z_i \sum_j s_{ij} z_j + \sum_i P_i z_i^2\right\}\theta^{(2\nu)},$$

d'où

$$\theta^{(2\nu+2)} = (2\nu+1)\theta^{(2)}\theta^{(2\nu)},$$

et, par suite,

$$\theta^{(2\nu)} = (2\nu-1)(2\nu-3)\ldots 3.1\,\sigma^{2\nu}.$$

Application à la convergence de $A_h^{(n)}$ et de $f_{hk}^{(n)}$. — Posons

$$u_n = \mathfrak{M}[A_h^{(n)} - M]^4 = \frac{\mathfrak{N}_{hn}^{(4)}}{n^4}.$$

On a vu que, dans le cas régulier, $n^2 u_n$ converge vers une limite $\theta^{(4)}$. Dès lors la série $\sum u_n$ est convergente ([1]). Or, on sait (voir par exemple le premier Livre de cet Ouvrage, p. 238) que cela suffit pour établir la convergence « presque certaine » de $A_h^{(n)}$ vers M.

Ainsi, non seulement la valeur moyenne (au sens du Calcul des Probabilités) $\mathfrak{M}A_h^{(n)}$ de $A_h^{(n)}$ converge toujours vers une limite M_h, *mais encore, il y a, dans le cas régulier* (où $M_h = M$) ([1]), *une probabilité égale à l'unité que la moyenne arithmétique aléatoire* $A_h^{(n)} = \dfrac{Y_h^{(1)} + \ldots + y_h^{(n)}}{n}$ *converge vers le nombre certain* M.

En particulier, il est ainsi démontré qu'*au moins dans le cas régulier, la fréquence (aléatoire)* $f_{hk}^{(n)}$ *de l'état* E_k *au cours de n épreuves à partir de l'état* E_h *converge presque certainement vers une limite certaine (à savoir* P_k, *la limite de* $\mathfrak{M}f_{hk}^{(n)}$).

M. Kolmogoroff qui a attiré mon attention sur ces points, a bien voulu m'informer qu'ils peuvent être ainsi complétés dans le cas singulier, comme

([1]) Si l'on admet ce résultat de la page 153, que, dans le cas semi-régulier, $\dfrac{\mathfrak{N}_{hn}^{(3)}}{n}$ est borné, il en sera de même de $f_k^{(4)}(n)$ et d'après la note C, page 279, il en résultera que $\dfrac{\mathfrak{N}_{hn}^{(4)}}{n}$ est aussi borné et, par suite, que Σu_n reste convergente dans le cas semi-régulier. Les deux conséquences qui suivent en italiques restent donc valables dans le cas semi-régulier.

dans le cas régulier, non seulement la valeur moyenne $\mathfrak{M} f_{hk}^{(n)}$ de $f_{hk}^{(n)}$ converge comme on l'a vu, page 141, vers une limite π_{hk}, mais la fréquence aléatoire $f_{hk}^{(n)}$, elle-même, converge presque certainement vers une limite. Cette limite, qu'on peut désigner par f_{hk}, est, en général, aléatoire et sa valeur moyenne $\mathfrak{M} f_{hk}$ est égale à π_{hk}. Dans le cas semi-régulier, nous avons établi, page 152, que $f_{hk}^{(n)}$ converge « en probabilité » vers π_k, donc, dans ce cas, f_{hk} est un nombre certain égal à π_k.

D'après M. Kolmogoroff, ces résultats peuvent être déduits d'un théorème général ergodique de Birkhoff-Khintchine, une démonstration élémentaire, non encore publiée, de ces résultats a été aussi donnée par M. Barkalaja.

Ultérieurement, M. Dœblin m'en a communiqué une démonstration reposant sur la notion de groupement final introduite plus loin, page 174 et prouvant en outre que si f_{hk} est aléatoire, cette limite ne prend, presque certainement, que deux valeurs, dont l'une est zéro.

Limite de la loi réduite de répartition. — Les résultats précédents vont être précisés dans le cas régulier. Les expressions des moments obtenues plus haut montrent qu'on a

$$\lim_{n \to \infty} \frac{\mathfrak{M}_{k,n}^{(2\nu)}}{n^{\nu}} = (2\nu - 1) \ldots 3.1.\sigma^{2\nu} = (2\nu - 1) \ldots 3.1 \lim_{n \to \infty} \left[\frac{\mathfrak{M}_{k,n}^{(2)}}{n}\right]^{\nu},$$

ou, *si σ n'est pas nul*,

$$\lim_{n \to \infty} \frac{\mathfrak{M}_{k,n}^{(2\nu)}}{[2\mathfrak{M}_{k,n}^{(2)}]^{\nu}} = \frac{(2\nu - 1) \ldots 3.1}{2^{\nu}}.$$

C'est-à-dire que, pour t variant par valeurs paires, $\dfrac{\mathfrak{M}_{k,n}^{(t)}}{\left[\sqrt{2\mathfrak{M}_{k,n}^{(2)}}\right]^{t}}$ a une limite. Quand t est impair, on a à considérer

$$\frac{\mathfrak{M}_{k,n}^{(2\nu+1)}}{\left[\sqrt{2\mathfrak{M}_{k,n}^{(2)}}\right]^{2\nu+1}} = \frac{n^{\nu}\,\theta_k^{(2\nu+1)} + \ldots}{\left[\sqrt{2n\sigma^2 + \ldots}\right]^{2\nu+1}},$$

qui est équivalent à

$$\frac{\theta_k^{(2\nu+1)}}{\sqrt{n}(\sigma\sqrt{2})^{2\nu+1}},$$

et, par suite, tend vers zéro.

Les limites respectives des quantités

$$\mu_{h,n}^{(t)} = \frac{\mathfrak{M}_{h,n}^{(t)}}{\left[\sqrt{2\mathfrak{M}_{h,n}^{(2)}}\right]^{t}}$$

quand n croît, sont donc $\frac{(2\nu-1)\ldots 3\, 1}{2^\nu}$ quand t est pair, et zéro quand t est impair. Elles sont donc égales aux moments d'une variable aléatoire Y suivant « la seconde loi de Laplace », c'est-à-dire telle que

$$\{\text{Probab } Y < y\} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{y} e^{-t^2}\,dt$$

Or,

$$\mu_{h,n}^{(t)} = \mathfrak{M}_{E_h}\left\{\frac{(X_1 - M) + \ldots + (X_n - M)}{\sqrt{2\,\mathfrak{M}_{h,n}^{(2)}}}\right\}^t$$

$$= \mathfrak{M}\left\{\frac{(Y_h^{(1)} - M) + \ldots + (Y_h^{(n)} - M)}{\sqrt{2\,\mathfrak{M}_{h,n}^{(2)}}}\right\}^t.$$

En vertu du « second théoreme-limite de la Théorie des Probabilités (¹), — et toujours *en supposant* $\sigma \neq 0$, on voit que, dans le cas régulier, *la loi de répartition de la variable aléatoire*

$$Z_n = \frac{Y_h^{(1)} - M + \ldots + Y_h^{(n)} - M}{\sqrt{2\,\mathfrak{M}\,\{[Y_h^{(1)} - M] + \ldots + [Y_h^{(n)} - M]\}^2}}$$

converge vers la seconde loi de Laplace, c'est-à-dire que

$$\lim_{n \to \infty} \text{Probab. } \{Z_n < x\} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-t^2}\,dt.$$

D'ailleurs, les moments de Z_n ont les mêmes limites respectives que ceux de

$$T_n = \frac{Y_h^{(1)} - M + \ldots + Y_h^{(n)} - M}{\sigma\sqrt{2n}} = \frac{A_h^{(n)} - M}{\sigma\sqrt{\frac{2}{n}}}.$$

Ainsi, dans le cas régulier (²), *la loi de répartition de* $\frac{A_h^{(n)} - M}{\sigma\sqrt{\frac{2}{n}}}$ *tend*

(¹) *Voir*, par exemple, ce Traité, I, 2, p. 67 et 70, ou, pour une démonstration nouvelle, Fréchet et Shohat [141].

(²) MM. Onicescu [4] et Mihoc publient au moment de la correction des épreuves de notre livre l'esqnisse d'une nouvelle démonstration du même résultat. Se plaçant dans un cas un peu plus général (le cas semi-régulier en supposant $\sigma_h \neq 0$), ils opèrent par la méthode des fonctions caractéristiques (définies au premier Livre, p. 106).

aussi vers la seconde loi de Laplace, pourvu, bien entendu, qu'on ait $\sigma \neq 0$, c'est-à-dire (page 153), *pourvu que la valeur moyenne du carré de* $\{Y_h^{(1)} - M + \ldots + Y_h^{(n)} - M\}$ *croisse indéfiniment avec* n.

Rappelons ici que $Y_h^{(n)}$ est la valeur prise par la variable aléatoire X à une $n^{\text{ième}}$ épreuve quand le système était initialement à l'état E_h. On voit que, dans le cas régulier, la loi réduite limite de $A_h^{(n)}$ est indépendante de l'état initial E_h.

Précisions sur les moments d'ordre supérieur — Les raisonnements qui ont permis d'établir la limite de la loi réduite de répartition ne nécessitent que la démonstration de l'existence des θ d'ordre impair et le calcul des θ d'ordre pair Mais la méthode employée permet d'aller plus loin quand on reste dans le cas régulier. On a formé plus haut le développement de $\mathfrak{M}_{k,n}^{(1)}$. Procédons de même pour $\mathfrak{M}_{k,n}^{(2)}$. On a déjà vu que dans (50), $B_{k2} = \sigma^2$ Calculons maintenant A_{k2}.

En remplaçant dans (39) et (39 *bis*), pour $\nu = 2$, $\mathfrak{M}_{k,n}^{(2)}$ et $\mathfrak{M}_{k,n}^{(1)}$ par leurs expressions (40) et (50), on obtient

$$A_{k2} - \sum_i p_{ki} A_{i2} = 2 \sum_i p_{ki} z_i \sum_j s_{ij} z_j + \sum_i p_{ki} z_i^2 - \sigma^2 = F_k$$

En réalité, on devrait ajouter au second membre une somme de termes dont on sait qu'ils tendent vers zéro avec $\frac{1}{n}$. Mais en faisant croître n indéfiniment, on peut supprimer ce terme additionnel. Dès lors les A_{k2} sont des solutions constantes du système

$$y_k(n) - \sum_l p_{kl} y_l(n-1) = F_k.$$

D'après la Note C, page 279, appliquée à ce système, on a

$$A_{k,2} = F_k + \sum_l s_{kl} F_l + A,$$

où A est indépendant de k, d'où

$$A_{k2} - A = 2\sum_i p_{ki} z_i \sum_j s_{ij} z_j + \sum_i p_{ki} z_i^2 - \sigma^2$$

$$+\sum_i \left[\sum_l s_{kl} p_{li}\right] \left\{ 2 z_i \sum_j s_{ij} z_j + z_i^2 \right\}$$

$$= 2\sum_i \sum_j p_{ki} s_{ij} z_i z_j + \sum_i p_{ki} z_i^2 - \sigma^2$$

$$+\sum_i (s_{ki} - p_{ki} + P_i)\left[2 z_i \sum_j s_{ij} z_j + z_i^2 \right]$$

$$= 2\sum_i s_{ki} z_i \sum_j s_{ij} z_j + \sum_i s_{ki} z_i^2$$

Ainsi,

$$\mathfrak{M}_{k,n}^{(2)} = n\sigma^2 + 2\sum_i s_{ki} z_i \sum_j s_{ij} z_j + \sum_i s_{ki} z_i^2 + A + \varphi_{k,2}^{(n)}, \tag{53}$$

où $\varphi_{k,2}^{(n)}$ converge vers zéro avec $\frac{1}{n}$.

Le développement (53) va nous servir pour calculer $\theta_i^{(3)}$. Avec les notations de la page 154, on a

$$f_{kn}^{(3)} = n B'_k + A'_k + \eta_k(n)$$

D'où (Note C, p. 279), puisque $\sum_k P_k B'_k = 0$,

$$\theta_i^{(3)} = B_{i,3} = B'_i + \sum_k s_{ik} B'_k + \sum_k P_k A'_k$$

avec

$$B'_k = 3\sum_j p_{kj} z_j \sigma^2,$$

$$A'_k = 3\sum_j p_{kj} z_j A_{j2} + 3\sum_j p_{kj} z_j^2 \sum_i s_{ji} z_i + \sum_i p_{ki} z_i^3.$$

$$\theta_i^{(3)} = 3\sum_j \left[p_{ij} + \sum_k s_{ik}\, p_{kj} \right] z_j \sigma^2 + Q$$

$$= 3\sum_j (s_{ij} + P_j)\, z_j \sigma^2 + Q'$$

$$= 3\sum_j s_{ij}\, z_j \sigma^2 + Q',$$

et enfin

$$\theta_i^{(3)} = 3[\sigma^2 \theta_i^{(1)} + Q], \tag{54}$$

où $Q = \frac{1}{3}\sum_k P_k A'_k$ est une quantité indépendante de i.

On peut calculer explicitement Q

$$3Q = \sum_k P_k A'_k = \sum_i \left(\sum_k P_k\, p_{ki} \right) \{ 3 z_i A_{i,2} + 3 z_i^2 \theta_i^{(1)} + z_i^3 \} = \sum_i P_i \{ \quad \},$$

avec

$$3\sum_i P_i z_i A_{i,2} = 3\sum_i (P_i z_i) \left[A + \sum_j s_{ij} z_j^2 + 2\sum_j \sum_l s_{ij} s_{jl} z_j z_l \right]$$

et

$$\sum_i P_i z_i A = 0,$$

de sorte que

$$3Q = \sum_i P_i z_i^3 + 3\sum_i \sum_j P_i s_{ij} z_j z_i^2 \tag{55}$$

$$+ 3\sum_i \sum_j P_i s_{ij} z_i z_j^2 + 6\sum_i \sum_j \sum_l P_i s_{ij} s_{jl} z_i z_j z .$$

Nous allons maintenant étendre la formule (54) aux $\theta_i^{(2\nu+1)}$.

Calcul de $\theta_i^{(2\nu+1)}$. — Cherchons d'abord un terme de plus du développement de

$$\mathfrak{N}_{k,n}^{(2\nu)} = n^\nu \theta^{(2\nu)} + n^{\nu-1} J_{k,2\nu} + \ldots .$$

On aura, en remplaçant dans l'équation analogue à (39)

$$n^\nu \theta^{(2\nu)} + n^{\nu-1} J_{k,2\nu} + \quad - \sum_l p_{kl}[(n-1)^\nu \theta^{(2\nu)} + (n-1)^{\nu-1} J_{l,2\nu} + \ldots]$$

$$= f_k^{(2\nu)}(n-1) = \quad C_{2\nu}^1 \sum_l p_{kl} z_l [(n-1)^{\nu-1} \theta_l^{(2\nu-1)} + \ .\ ]$$

$$+ C_{2\nu}^2 \sum_l p_{kl} z_l^2 [(n-1)^{\nu-1} \theta^{(2\nu-2)} + \quad] + \ .$$

D'où, en égalant les coefficients de $n^{\nu-1}$

$$J_{k,2\nu} - \sum_l p_{kl} J_{l,2\nu}$$

$$= C_{2\nu}^1 \sum_l p_{kl} z_l \theta_l^{(2\nu-1)} + C_{2\nu}^2 \left(\sum_l p_{kl} z_l^2 \right) \theta^{(2\nu-2)} - \nu \theta^{(2\nu)} = T_l$$

Alors, d'après la Note C, page 279,

$$J_{k,2\nu} = R^{(2\nu)} + T_k + \sum_l s_{kl} T_l,$$

où $R^{(2\nu)}$ est indépendant de k et

$$T_k + \sum_j s_{kj} T_j = \sum_l \left\{ \left[p_{kl} + \sum_j s_{kj} p_{jl} \right] [C_{2\nu}^1 z_l \theta_l^{(2\nu-1)} + C_{2\nu}^2 z_l^2 \theta^{(2\nu-2)}] \right\} - \nu\theta^{(2\nu)}$$

$$= C_{2\nu}^1 \sum_l (s_{kl} + P_l) z_l \theta_l^{(2\nu-1)} + C_{2\nu}^2 \sum_l (s_{kl} + P_l) z_l^2 \theta^{(2\nu-2)} - \nu\theta^{(2\nu)}$$

D'où

$$J_{k,2\nu} = K^{(2\nu)} + H_k^{(2\nu)},$$

où $K^{(2\nu)}$ est indépendant de k et

$$H_k^{(2\nu)} = C_{2\nu}^1 \sum_l s_{kl} z_l \theta_l^{(2\nu-1)} + C_{2\nu}^2 \left[\sum_l s_{kl} z_l^2 \right] \theta^{(2\nu-2)}.$$

Ceci va nous permettre de calculer $\theta_l^{(2\nu-1)}$. On a vu, page 157, l'expression de $f_k^{(2\nu+1)}(n)$.

On sait (Note C, p 279), qu'alors, puisque $\sum_l P_l F_l = 0$, on a

$$\theta_k^{(2\nu+1)} = F_k + \sum_j s_{kj} F_j + \frac{1}{\nu} \sum_j P_j G_j = E_k^{(2\nu+1)} + D^{(2\nu+1)},$$

avec

$$E_h^{(2\nu+1)} = C_{2\nu+1}^1\,\theta^{(2\nu)} \sum_i \Big[p_{hi} + \sum_j s_{hj}\, p_{ji} \Big] z_i$$

$$= C_{2\nu+1}^1\,\theta^{(2\nu)} \sum_i (s_{hi} + P_i)\, z_i = (2\nu+1)\,\theta_h^{(1)}\,\theta^{(2\nu)},$$

d'où

$$\theta_h^{(2\nu+1)} = D^{(2\nu+1)} + (2\nu+1)\,\theta_h^{(1)}\,\theta^{(2\nu)}, \tag{56}$$

et

$$\nu D^{(2\nu+1)} = \sum_i \Big(\sum_j P_j\, p_{ji} \Big) \{ C_{2\nu+1}^1\, z_i\, J_{i,2\nu} + \quad \}$$

$$= C_{2\nu+1}^1 \Big(\sum_i P_i z_i \Big) K^{(2\nu)} + C_{2\nu+1}^1 \sum_i P_i z_i H_i^{(2\nu)}$$

$$+ C_{2\nu+1}^2 \sum_i P_i z_i^2\, \theta_i^{(2\nu-1)} + C_{2\nu+1}^3 \Big(\sum_i P_i z_i^3 \Big) \theta^{(2\nu-2)}$$

Le premier terme est nul; le second est

$$C_{2\nu+1}^1 \sum_i P_i z_i \Big\{ C_{2\nu}^1 \sum_j s_{ij}\, z_j\, \theta_j^{(2\nu-1)} + C_{2\nu}^2 \sum_j s_{ij}\, z_j^2\, \theta^{(2\nu-2)} \Big\}.$$

D'où

$$\nu D^{(2\nu+1)} = \frac{(2\nu+1)(2\nu)}{2} \Big\{ \sum_j \Big[P_j z_j^2 + 2 \sum_i P_i s_{ij}\, z_i z_j \Big] \theta_j^{(2\nu-1)} \Big\}$$

$$+ \frac{2\nu(2\nu+1)(2\nu-1)}{6} \Big\{ \sum_i P_i z_i^3 + 3 \sum_i \sum_j P_i s_{ij}\, z_i z_j^2 \Big\} \theta^{(2\nu-2)},$$

et en remplaçant $\theta_j^{(2\nu-1)}$ par $D^{(2\nu-1)} + (2\nu-1)\,\theta_j^{(1)}\,\theta^{(2\nu-2)}$

$$D^{(2\nu+1)} - (2\nu+1)\, D^{(2\nu-1)}\,\sigma^2$$

$$= \frac{(2\nu+1)(2\nu-1)}{3}\,\theta^{(2\nu-2)} \Big\{ \sum_i P_i z_i^3 + 3 \sum_i \sum_j P_i s_{ij}\, z_i z_j^2 + 3 \sum_i \sum_j P_i s_{ij}\, z_i^2 z_j$$

$$+ 6 \sum_i \sum_j \sum_l P_i s_{ij}\, s_{jl}\, z_i z_j z_l \Big\},$$

$$D^{(2\nu+1)} - (2\nu+1) D^{(2\nu-1)}\,\sigma^2 = \frac{(2\nu+1)(2\nu-1)}{3}\,\theta^{(2\nu-2)}\, 3Q \tag{57}$$

$$= (2\nu+1)(2\nu-1)(2\nu-3)\ldots 7.5.3\,\sigma^{2\nu-2}\, Q.$$

En transformant (57) au moyen de (56), on trouve

$$\theta_k^{(2\nu+1)} - (2\nu+1)\theta_k^{(2\nu-1)}\sigma^2 = (2\nu+1)(2\nu-1)\ldots 7.5.3\,\sigma^{2\nu-2}Q$$

On en tire par récurrence

$$\theta_k^{(2\nu+1)} = (2\nu+1)\theta^{(2\nu)}\left[\theta_k^{(1)} + \nu\frac{Q}{\sigma^2}\right],$$

avec, d'après (54)

$$\theta_k^{(3)} = 3\sigma^2\left(\theta_k^{(1)} + \frac{Q}{\sigma^2}\right),$$

d'où *la formule cherchée*

$$(58)\qquad \theta_k^{(2\nu+1)} = (2\nu+1)(2\nu-1)\ldots 5.3\,\sigma^{2\nu-2}[\sigma^2\theta_k^{(1)} + \nu Q],$$

où $\theta_k^{(1)} = \sum_l s_{kl}z_l$ et où Q est donné par la formule (55)

d — Étude du cas positivement régulier

Remarque — Nous avons observé plus haut, page 68, que lorsqu'on s'intéresse aux variables aléatoires en chaîne, le cas positivement régulier n'a généralement rien de plus intéressant que le cas régulier qui reste le seul important.

Cependant quand on s'intéresse plus spécialement au comportement des probabilités, le fait qu'un des P_k soit nul peut avoir de l'importance, et il vaut la peine de distinguer le cas positivement régulier

On peut d'abord considérer l'hypothèse $P_k = 0$ comme un cas particulier de l'hypothèse $\Pi_{jk} = 0$.

Condition pour qu'un, au moins, des Π_{jk} soit nul — On a vu que

$$\Pi_{jh} = \sum_i p_{ji}\Pi_{ih} = \sum_i \Pi_{ji}p_{ih}, \qquad \sum_i \Pi_{ji} = 1.$$

Supposons que Π_{jh} soit nul; alors l'ensemble a des indices k' tels que $\Pi_{jk'} = 0$ contient au moins un élément, à savoir h. D'ailleurs $\sum_k \Pi_{jk} = 1$; donc l'ensemble b des indices k restant, c'est-à-dire des

indices k'' tels que $\Pi_{jk''} \neq 0$ contient aussi un élément au moins. On a donc

$$0 = \Pi_{jk'} = \sum_{k''} \Pi_{jk''} p_{k''k'},$$

avec

$$\Pi_{jk''} \neq 0 \qquad \text{et} \qquad \Pi_{jk''} p_{k''k'} \geqq 0$$

D'où $p_{k''k'} = 0$ quand k'', k' appartiennent respectivement à b et a

Par conséquent, le tableau D des p_{jk} est *décomposable*, entendant par là qu'on peut décomposer les indices 1, 2, ..., r en deux groupes disjoints a, b, chacun contenant au moins un élément et tels que pour k', k'' pris arbitrairement sur a et b, on ait $p_{k''k'} = 0$.

(Dans ce cas, on peut ranger les états E_j de sorte que D prenne la forme

où le rectangle R est formé uniquement de zéros.)

Réciproquement, si D est décomposable au sens indiqué, comme on a

$$P^{(2)}_{k''k'} = \sum_i p_{k''i} p_{ik'},$$

on aura $P^{(2)}_{k''k'} = 0$, car si i appartient à a, $p_{k''i} = 0$ et si i appartient à b, $p_{ik'} = 0$. En général, si $P^{(s)}_{k''k'} = 0$ pour $s < n$, on aura

$$P^{(n)}_{k''k'} = \sum_i p_{k''i} P^{(n-1)}_{ik'} = 0,$$

pour la même raison. Ainsi on voit que dans la matrice $D^{(n)}$ des $P^{(n)}_{ik}$ le rectangle correspondant à R est aussi formé de zéros. Finalement,

si D est décomposable, il en est de même, quel que soit n, pour le tableau $D^{(n)}$ des $P_{jh}^{(n)}$. Il en résulte aussi que $P_{k''k'}^{(n)}$ étant nul quel que soit n, non seulement la limite en moyenne $\Pi_{k''k'}$ de $P_{k''k'}^{(n)}$ est nulle, mais même la limite au sens ordinaire de $P_{k''k'}^{(n)}$ existe aussi et est nulle.

En résumé, la *condition nécessaire et suffisante pour que l'une au moins des limites en moyenne* Π_{jh} *de* $P_{jh}^{(n)}$ *soit nulle est que le tableau* D des p_{jh} *soit décomposable*. Et alors le tableau des Π_{jh} est aussi décomposable ainsi que les tableaux $D^{(n)}$ des $P_{jh}^{(n)}$. Et pour au moins un couple k', k'', la suite de $P_{k''k'}^{(n)}$ converge vers zéro non seulement en moyenne, mais au sens ordinaire. Il en résulte, en particulier, que :

1° Dans le cas positivement régulier, D et chacun des $D^{(n)}$ sont nécessairement indécomposables ;

2° Si D ou l'un des $D^{(n)}$ est indécomposable, le cas régulier ne peut être que positivement régulier.

Observons d'ailleurs que si D est indécomposable, il n'en résulte pas que les $D^{(n)}$ le soient, comme le montre le cas où r étant égal à 2, le tableau de p_{jk} est $\begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}$, car alors celui de $P_{jk}^{(2)}$ est $\begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix}$.

Condition pour le cas positivement régulier. — Dans le cas positivement régulier, puisque chaque $P_{jh}^{(n)}$ converge vers une limite $\neq 0$, $D^{(n)}$ a ses termes positifs à partir d'un certain rang. A fortiori, on peut choisir n assez grand pour que : 1° $D^{(n)}$ soit indécomposable ; 2° la diagonale de $D^{(n)}$ ne soit pas toute nulle.

Réciproquement, supposons ces deux conditions réalisées, qui peuvent l'être pour une valeur ν de n sans que $D^{(\nu)}$ ait ses termes tous positifs. Il existe au moins un indice l tel que $P_{ll}^{(\nu)} \neq 0$. Or, on a

$$P_{jl}^{(m+1)\nu} \geqq P_{jl}^{(m\nu)} P_{ll}^{(\nu)}.$$

Pour un indice j déterminé, si, pour une valeur de n, $P_{jl}^{(n\nu)}$ est $\neq 0$, il en sera donc de même pour les valeurs de n supérieures. Or, $D^{(\nu)}$ étant indécomposable, en raisonnant sur les $P_{jl}^{(n\nu)}$ comme plus haut sur les $P_{jl}^{(n)}$, on voit que les $P_{jl}^{(n\nu)}$ ne peuvent tendre vers zéro quand n croît ; ils sont donc $\neq 0$ pour au moins une valeur de n et, par suite, à partir d'une certaine valeur de n, soit N_j. Si M est le plus grand des nombres $N_1, \ldots, N_r$, on voit qu'il existe un entier $\alpha = M\nu$ tel que

dans $D^{(\alpha)}$ une ligne tout entière soit $\neq 0$. Cela suffit, comme nous l'avons déjà observé (p. 31) pour qu'on soit dans le cas régulier Et comme $D^{(M\nu)}$ est indécomposable. on est bien dans les cas positivement régulier. En résumé, nous avons donné une nouvelle démonstration d'un théorème dû à M. de Mises [1] : *pour qu'on soit dans le cas positivement régulier, il faut et il suffit qu'il existe un entier* ν *tel que* $D^{(\nu)}$ *soit indécomposable et que sa diagonale principale ne soit pas toute nulle.*

Remarque. — Il est bon d'observer que *le cas régulier peut se présenter même quand le tableau* D *des* p_{ik} *est décomposable* Tel est le cas où D est le tableau

$$\left\| \begin{matrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} \end{matrix} \right\|.$$

On voit facilement qu'alors les éléments de $D^{(n)}$ convergent vers ceux du tableau

$$\left\| \begin{matrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{matrix} \right\|,$$

où les éléments d'une même ligne sont égaux entre eux. On est bien dans le cas régulier (*non positivement*).

Une propriété des racines de modules 1 de l'équation en s. — Appelons, comme à la page 105, équation en s, l'équation $\Delta(s) = 0$

Les racines de module 1 de cette équation vérifient une même équation binome $s^N = 1$.

Il est clair que pour toute racine s de l'équation en s, les équations

$$(59) \qquad s w_k = \sum_j p_{kj} w_j \qquad (k = 1, 2, \ldots, r)$$

admettent au moins un système de solutions $W_1, \ldots, W_r$ non toutes

nulles. En multipliant par s, en remplaçant au second membre les $s w_j$ par leurs expressions et en répétant cette opération, on voit aussi que ce système de solutions vérifiera le système

$$s^{N} w_k = \sum_j P_{kj}^{(N)} w_j \qquad (k = 1, 2, \ldots, r) \tag{60}$$

Si les W_j sont égaux, comme leur valeur commune est $\neq 0$, on aurait, d'après (59), $s = 1$. Supposons $|s| = 1$, mais $s \neq 1$. Alors les W_j ne sont pas tous égaux et (puisque $s^N = 1$) vérifient le système

$$w_k = \sum_j P_{kj}^{(N)} w_j \qquad (k = 1, 2, \ldots, r) \tag{61}$$

Procédons comme dans la Note (¹) de la page 107.

Soit $R = |W_h|$ le plus grand module des W_j; on aura $R \neq 0$ et il existe certainement au moins un $W_j \neq W_h$. Appelons α l'ensemble des indices j' tels que $W_{j'} \neq W_h$ et β l'ensemble des j'' tels que $W_{j''} = W_h$. Chacun de ces ensembles contient au moins un indice, et l'on a

$$W_{j''} \sum_j P_{j''j}^{(N)} = W_{j''} = \sum_j P_{j''j}^{(N)} W_j,$$

d'où, puisque $|W_{j''}| = R \neq 0$,

$$\sum_j P_{j''j}^{(N)} \left[1 - \frac{W_j}{W_{j''}}\right] = 0.$$

D'où, en posant $\frac{W_j}{W_{j''}} = u_j + i v_j$,

$$\sum_j P_{j''j}^{(N)} [1 - u_j] = 0. \tag{62}$$

Or,

$$0 \leqq (v_j)^2 \leqq 1 - (u_j)^2,$$

d'où $[1 - u_j] \geqq 0$ et l'on n'a $1 - u_j = 0$ que si $v_j = 0$, d'où $W_j = W_{j''}$, c'est-à-dire que $1 - u_j > 0$ quand j est un des j' de l'ensemble α et $1 - u_j = 0$ quand j appartient à β. Dès lors, en vertu de (62), on a nécessairement $P_{j''j'}^{(N)} = 0$. Autrement dit, $D^{(N)}$ est décomposable.

En résumé, *quand l'équation en s possède une racine différente*

de 1 *et de module* 1, *l'un au moins des* $D^{(n)}$ *de rang supérieur à* 1 *est décomposable.*

Conséquences — Nous savons que, pour que les $P_{jk}^{(n)}$ convergent au sens ordinaire, il faut et il suffit que l'équation en s n'ait pas de racine de module 1 autre que l'unité. Donc :

Pour que les $P_{jk}^{(n)}$ *convergent au sens ordinaire, il suffit que tous les* $D^{(n)}$ *de rang supérieur à* 1 *soient indécomposables.* L'exemple donné ci-dessus, page 170, montre que cette condition n'est pas nécessaire.

Le raisonnement qui vient d'être exposé pour les équations (60) peut aussi se répéter sur les équations

$$w_k = \sum_j p_{kj} w_j, \tag{63}$$

et conduit au résultat suivant : si ce système d'équations admet un système de solutions non toutes égales, alors D est décomposable Dès lors, si D est indécomposable, ce système ne peut admettre que des solutions égales. Or, les limites en moyenne, Π_{jk} des $P_{jk}^{(n)}$, vérifient les équations

$$\Pi_{ki} = \sum_j p_{kj} \Pi_{ji},$$

de sorte que le système (63) admet le système de solutions

$$w_1 = \Pi_{1i}, \qquad \ldots, \qquad w_r = \Pi_{ri}$$

Donc, si D est indécomposable, on a $\Pi_{1i} = \Pi_{2i} = \ldots = \Pi_{ri}$, quel que soit i.

En résumé, *quand* D *est indécomposable, les limites en moyenne* Π_{ki} *des* $P_{ki}^{(n)}$ *sont indépendantes du premier indice.* Nous avons vu aussi plus haut qu'elles sont, dans le cas actuel, toutes différentes de zéro.

Or, pour que les Π_{ki} soient indépendantes du premier indice, il faut et il suffit que l'unité ne soit pas racine multiple de l'équation en s. On retrouve ainsi, comme conséquence des deux théorèmes précédents, un résultat démontré, en suivant une autre méthode,

par M. Romanovsky : si D n'est pas décomposable, l'unité ne peut être racine multiple de l'équation en s.

Signalons aussi une autre conséquence curieuse des résultats précédents : *si les limites* Π_{jk} (en moyenne ou, *a fortiori*, ordinaires) *des* $P_{jk}^{(n)}$ *sont toutes* $\neq 0$, *alors elles sont nécessairement indépendantes du premier indice.*

On pourrait appeler *cas positif* le cas où les Π_{jk} sont tous > 0. On voit que : *si* D *est indécomposable, on est à la fois dans le cas positif et dans le cas semi-régulier.*

Nouvelle forme de la condition pour le cas positivement régulier. — D'après ce qui précède : quand D et tous ses itérés $D^{(n)}$ sont indécomposables, on est nécessairement dans le cas positivement régulier. D'ailleurs la réciproque est vraie. En effet, supposons que $D^{(n)}$ soit décomposable pour au moins une valeur ν de n, avec $\nu \geqq 1$. Alors il y a au moins un couple k', k'' tel que $P_{k''k'}^{(n\nu)} = 0$ quel que soit n. Donc : ou bien on n'est pas dans le cas régulier, ou bien si l'on est dans le cas régulier, la limite, qui est nulle, de $P_{k''k'}^{(n\nu)}$, quand n croît, est égale à la limite $P_{k'}$ de $P_{k''}^{(t)}{}_{k'}$ quand t croît par valeurs entières successives. Par suite $P_{k'}$ devrait être égal à zero et l'on serait dans le cas régulier non positivement.

En résumé, *pour qu'on soit dans le cas positivement régulier, il faut et il suffit que le tableau* $D^{(n)}$ *des* $P_{jk}^{(n)}$ *soit indécomposable quelle que soit la valeur entière* (*supérieure ou égale à* 1) *de* n.

e. — Répartition des états possibles en groupements indécomposables en une épreuve

Interprétation des tableaux décomposables. — Nous avons donné une définition formelle des tableaux ou matrices décomposables. Il n'est pas sans intérêt de donner aussi à cette définition une forme plus concrète. Soit A, l'ensemble des états $E_{k'}$, B celui des $E_{k''}$ définis page 168 : la condition $p_{k''k'} = 0$ exprime qu'on peut décomposer l'ensemble G des états possibles en deux ensembles A, B, comprenant chacun un état au moins et tels qu'il soit impossible de passer en une épreuve d'un état de B à un état de A (ou plus précisément tel que

la probabilité d'un tel passage soit nulle). Quand il en est ainsi, nous dirons que l'ensemble G des états possibles est décomposable et nous sous-entendrons : en une épreuve. On peut dire avec M. de Mises que la sortie de B est impossible (ou plutôt « presque impossible », c'est-à-dire de probabilité nulle). Observons bien qu'ici A et B ne jouent, en général, pas le même rôle.

Décomposition en groupements indécomposables. — Parmi les décompositions possibles, s'il en existe au moins une, $G = A + B$, de G en deux ensembles A, B de la nature indiquée, il en est au moins une, soit $G = \alpha + \beta$, qui comporte pour α le plus petit nombre ($\geqq 1$) d'états possibles.

Si la matrice des p_{jh} correspondant à α était décomposable, les états de α pourraient être répartis en deux groupements α_1, α_2 tels que $p_{jh} = 0$, si E_j, E_h appartiennent respectivement à α_2 et α_1. Dès lors, on aurait aussi $p_{jh} = 0$, si E_h continuant à appartenir à α_1, E_j appartenait à β ou à α_2. La décomposition dans les deux groupements $\beta' = \beta + \alpha_2$ et $\alpha' = \alpha_1$ fournirait pour α' un nombre de termes inférieur à celui des termes de α, ce qui est impossible.

Ainsi, quand l'ensemble G des états possibles est décomposable (en une épreuve), il peut être en particulier décomposé en deux groupements α, β sans élément commun, comprenant chacun un élément au moins, dont le premier, α, est indécomposable (en une épreuve) et du second desquels il est presque impossible de sortir.

La matrice D' des p_{ij} de β a ses termes $\geqq 0$ et les sommes de ses colonnes sont égales à l'unité. On peut donc raisonner sur D' comme sur D et par suite sur β comme sur G : ou bien β est indécomposable en une épreuve, ou bien on peut décomposer β en un groupement indécomposable en une épreuve et un autre groupement duquel il est presque impossible de sortir. Et ainsi de suite.

Finalement, on peut décomposer l'ensemble G *des états possibles en une suite ordonnée de groupements d'états :* g_1, g_2, ..., g_t, *deux à deux disjoints, comprenant chacun au moins un état de* G, *chacun indécomposable en une épreuve et tels qu'il soit presque impossible de passer de l'un d'eux à l'un des précédents.* (Bien entendu, quand G est indécomposable, le nombre de ces groupements indécomposables se réduit à un.)

Autrement dit, on peut ranger les états de G de sorte que D prenne une forme où un certain nombre de matrices carrées, indécomposables, de même ordre ou non, $A_1, \ldots, A_t$ étant disposées consécutivement à cheval sur la diagonale principale et de façon à l'occuper entièrement, tous les termes de D situés dans la région R (*fig.* 2) placée au-dessus de ces matrices est formée de zéros.

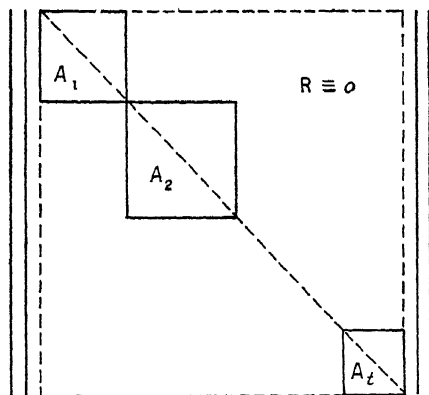



Fig. 2.

Les états se trouvent alors partagés en groupements $g_1, g_2, \ldots, g_t$, chaque groupement g_m correspondant aux états dont les indices se trouvent sur la diagonale principale de A_m. Et, d'après le procédé employé, $p_{i''i'} = 0$ quand i' appartient à $A_{m'}$, i'' à $A_{m''}$ avec $m' < m''$.

Il est visible que la somme de chaque colonne de A_t est égale à l'unité. Il peut arriver que parmi les matrices $A_1, A_2, \ldots, A_{t-1}$, il y en ait d'autres dont les sommes des colonnes sont toutes égales à l'unité. Alors les sommes des parties des colonnes de D qui sont au-dessous d'une telle matrice sont nécessairement nulles comme l'étaient déjà les parties qui sont au-dessus.

M. Romanovsky [3] qui a, le premier, formulé la décomposition ci-dessus dans le cas le plus général de Markoff, et l'a obtenue comme conséquence des travaux algébriques de Frobenius, avait donné aux deux sortes de groupements les noms de groupements isolés et de groupements de transition. Suivant une dénomination due à M. Dœblin (qui est arrivé, ainsi que M. Kolmogoroff et M. Fouillade, par une voie différente, à la décomposition précédente), nous

appellerons tout groupement correspondant à une telle matrice, un groupement *final* et les autres groupements, groupements *de passage*. On voit la raison de ces désignations.

Si g_m est un groupement final, il est presque impossible de passer d'un état de g_m à l'un quelconque des états n'appartenant pas à g_m La propriété reconnue plus haut à l'ordre adopté pour les groupements simples $g_1, g_2, \ldots, g_t$ ne sera donc pas troublée si nous

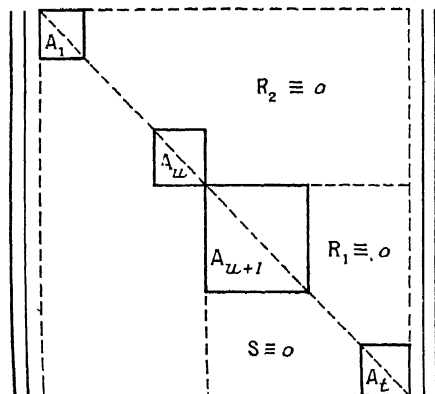



Fig. 3

reportons tous les groupements finals à la fin de cette suite. On obtiendra alors la disposition ci-contre (*fig.* 3).

La formule d'itération des $P_{jk}^{(n)}$ montre que dans $D^{(n)}$, les régions de la figure 3 correspondant à S, R_1, R_2 sont aussi formées de zéros.

Il peut ne pas y avoir de groupement de passage (en voir un exemple, p. 183), mais il y aura toujours au moins un groupement final g_t, qui peut se confondre avec l'ensemble G des états.

Si le Système appartient à un groupement final, il y restera presque sûrement. S'il appartient à un groupement de passage, il pourra y rester, mais s'il en sort, comme, à toute épreuve ultérieure, il est presque impossible qu'il rétrograde dans la suite des groupements $g_1, g_2, \ldots,$ il est presque certain qu'il ne pourra revenir à son groupement initial.

Quand i, k ne prennent que les valeurs relatives à un même groupement final g, on a, en représentant par $\sum_j^g$ une sommation faite en

donnant à j les valeurs des indices des états appartenant à g.

$$P_{ik}^{(n+m)} = \sum_{i}^{g} P_{ij}^{(n)} P_{jk}^{(m)}, \qquad \sum_{i}^{g} P_{ij} = 1 \qquad \text{avec} \quad p_{ij} \geqq 0.$$

De sorte que les formules relatives a G sont aussi valables quand on remplace G par g. On peut alors appliquer ce qui a été dit quand l'ensemble G des états possibles est indécomposable en une épreuve au cas où l'on remplace G par un de ses groupements finals (ce qui ne serait pas légitime pour un groupement de passage). Par exemple · a l'intérieur d'un groupement final les limites en moyenne $\Pi_{ij} = \Pi_j$ sont toutes $\neq 0$ et indépendantes du premier indice (quand l'état E_i correspondant à cet indice n'est pris que dans le même groupement final).

/ — Répartition des groupements finals en sous-groupements cycliques

Examen de $D^{(N)}$. — Puisque $P_{jk}^{(n)}$ est asymptotiquement périodique de période N. $P_{jk}^{(n+tN)}$ converge, quand t croît, vers une limite $Q_{jk}^{(n)}$ Et comme

$$P_{jk}^{(n+tN)} = \sum_{i} P_{ji}^{(n)} P_{ik}^{(tN)} = \sum_{i} P_{ji}^{(tN)} P_{ik}^{(n)},$$

on a, en posant $Q_{ji} = Q_{ji}^{(0)}$,

$$Q_{jk}^{(n)} = \sum_{i} P_{ji}^{(n)} Q_{ik} = \sum_{i} Q_{ji} P_{ik}^{(n)} \tag{61}$$

On peut considérer des événements en chaîne où le tableau des probabilités initiales de passage des p_{ik} est remplacé par celui des $P_{ik}^{(N)}$. Dans ce second problème, on sera dans un cas non oscillant, puisque les probabilités itérées seront les $P_{ik}^{(tN)}$ qui tendent vers des limites Q_{ik} quand t croît. On peut alors distinguer deux cas.

I. Dans le premier cas, les Q_{ik} sont tous $\neq 0$. Par suite, il existe au moins une valeur t_0 de t, assez grande pour que $P_{ik}^{(t_0 N)} \neq 0$ quels que soient i et k. On est donc (p. 31), dans le premier problème, comme dans le second, dans le cas positivement régulier.

II. Supposons maintenant que l'une des quantités Q_{ik} soit nulle. Alors, dans le second problème, la matrice des probabilités initiales est décomposable (d'après la page 164). On peut répartir les états en un certain nombre de groupements $C_1, \ldots, C_\rho$ qui sont indécomposables pour le second problème. Nous allons voir qu'on peut même aller plus loin en utilisant les formules de la page 110.

$$\Pi_{jk} = \sum_i P_{ji}^{(n)} \Pi_{ik} = \sum_i \Pi_{ji} P_{ik}^{(n)}.$$

Cas semi-régulier positif. — Supposons que dans le premier problème, on soit dans le cas semi-régulier positif (p. 169). Alors on aura

$$\Pi_k = \sum_{i=1}^{i=\nu} \Pi_i P_{ik}^{(N)}.$$

On a vu qu'on peut ranger $C_1, C_2, \ldots$ dans un ordre tel qu'il soit presque impossible au système de rétrograder sur cette suite dans le second problème.

Si E_k appartient à C_1, $P_{ik}^{(N)}$ est donc nul pour les E_i n'appartenant pas à C_1. On a donc

$$\Pi_k = \sum_i^{C_1} \Pi_i P_{ik}^{(N)},$$

d'où

$$\sum_i^{C_1} \Pi_i \sum_k^{C_1} P_{ik}^{(N)} = \sum_k^{C_1} \Pi_k = \sum_i^{C_1} \Pi_i,$$

et, par suite,

$$\sum_i^{C_1} \Pi_i \left[1 - \sum_k^{C_1} P_{ik}^{(N)} \right] = 0$$

Les crochets sont tous $\geqq 0$ et les Π_i sont tous > 0. Donc les crochets sont nuls, c'est-à-dire que, dans le second problème, le groupement C_1 est final. On aurait de même, si E_k appartient à C_2,

$$\Pi_k = \sum_i^{C_1} \Pi_i P_{ik}^{(N)} + \sum_i^{C_2} \Pi_i P_{ik}^{(N)},$$

$$\sum_i^{C_2} \Pi_i \left[1 - \sum_k^{C_2} P_{ik}^{(N)} \right] = \sum_i^{C_1} \Pi_i \left\{ \sum_k^{C_2} P_{ik}^{(N)} \right\}.$$

Puisque

$$\sum_{k}^{C_1} P^{(N)}_{ik} = 1 = \sum_{k=1}^{k=r} P^{(N)}_{ik},$$

quand E_i appartient à C_1, $\sum_{k}^{C_1} P^{(N)}_{ik}$ est alors nul, l'accolade ci-dessus est donc aussi nulle, et alors on voit, comme précédemment, que C_2 est aussi final. Et ainsi de suite : quand on est dans le cas semi-régulier positif, chacun des groupements C_1, C_2, ... est final dans le second problème. On a vu qu'alors les limites en moyenne des probabilités itérées (qui dans ce second probleme sont les limites Q_{jk} des $P^{(tN)}_{jk}$ quand t croît) sont indépendantes du premier indice et $\neq 0$, à l'intérieur de chaque groupement final Si donc E_j, E_k appartiennent respectivement à C_α, C_β, on a

$$0 = P^{(N)}_{jk} = P^{(2N)}_{jk} = \dots = P^{(tN)}_{jk} = \dots = Q_{jk}$$

quand $\alpha \neq \beta$, et

$$\lim_{t \to \infty} P^{(tN)}_{jk} = Q_{jk} = P_k \neq 0 \qquad \text{pour} \quad \alpha = \beta.$$

D'autre part, en vertu de (64), on a

$$Q^{(n)}_{jk} = \sum_{i=1}^{i=r} Q_{ji} P^{(n)}_{ik} = \sum_{i}^{C_\alpha} P_i P^{(n)}_{ik}, \tag{65}$$

donc $Q^{(n)}_{jk}$ a une valeur qui, sans être tout à fait indépendante de j, reste la même quand j varie dans C_α. On peut désigner cette valeur par $q^{(n)}_{\alpha k}$. Si dans la formule (64), on remplace n par $n + tN$ et fait croître t, on a

$$Q^{(n)}_{jk} = \sum_{i=1}^{i=r} Q^{(n)}_{ji} Q_{ik} = \sum_{i}^{C_\beta} Q^{(n)}_{ji} P_k,$$

ou

$$q^{(n)}_{\alpha k} = \sum_{i}^{C_\beta} q^{(n)}_{\alpha i} P_k = \lambda^{(n)}_{\alpha\beta} P_k,$$

avec

$$\lambda^{(n)}_{\alpha\beta} = \sum_{i}^{C_\beta} q^{(n)}_{\alpha i} \geqq 0. \tag{67}$$

On a, de plus,

$$1 = \lim_{t \to \infty} \sum_{h=1}^{h=l} P_{jh}^{(n+tN)} = \sum_{h=1}^{h=l} Q_{jh}^{(n)} = \sum_{\beta} \left\{ \lambda_{\alpha\beta}^{(n)} \left[\sum_{k}^{C_\beta} P_k \right] \right\},$$

avec

$$1 = \sum_{k=1}^{k=l} Q_{hk} = \sum_{k}^{C_\beta} P_k,$$

d'où

$$\sum_{\beta} \lambda_{\alpha\beta}^{(n)} = 1 \tag{68}$$

On va montrer que les λ sont nécessairement égaux à o ou 1.

On voit d'abord, comme pour (64), qu'on a

$$Q_{jk}^{(n+m)} = \sum_{i=1}^{i=l} Q_{ji}^{(n)} Q_{ik}^{(m)},$$

d'où, si E_i appartient à C_γ,

$$\lambda_{\alpha\beta}^{(n+m)} P_k = \sum_{\gamma} \lambda_{\alpha\gamma}^{(n)} \lambda_{\gamma\beta}^{(m)} \left[\sum_{i}^{C_\gamma} P_i \right] P_k,$$

et puisque $P_k \neq 0$ et $\displaystyle\sum_{i}^{C_\gamma} P_i = 1$,

$$\lambda_{\alpha\beta}^{(n+m)} = \sum_{\gamma} \lambda_{\alpha\gamma}^{(n)} \lambda_{\gamma\beta}^{(m)}. \tag{69}$$

Les relations (67), (68), (69) permettent de considérer les $\lambda_{\alpha\beta}$ comme les probabilités de passage en une épreuve de C_α en C_β dans un troisième problème où les états possibles seraient C_1, C_2, ...

Seulement, on a

$$\lambda_{\alpha\beta}^{(N)} P_k = Q_{jk} = \begin{cases} P_k & \text{pour } \alpha = \beta, \\ 0 & \text{pour } \alpha \neq \beta, \end{cases}$$

d'où

$$\lambda_{\alpha\beta}^{(N)} = \begin{cases} 1 & \text{si } \alpha = \beta, \\ 0 & \text{si } \alpha \neq \beta. \end{cases}$$

On a, pour $\beta \neq \alpha$,

$$0 = \lambda_{\alpha\beta}^{(N)} = \sum_{\gamma} \lambda_{\alpha\gamma} \lambda_{\gamma\beta}^{(N-1)} \quad \text{avec} \quad \lambda_{\alpha\gamma} = \lambda_{\alpha\gamma}^{(1)}. \tag{70}$$

Comme $\sum_\gamma \lambda_{\alpha\gamma} = 1$, il existe au moins un indice σ_1 tel que $\lambda_{\alpha\sigma_1} \neq 0$. Alors d'après (70) $\lambda_{\sigma_1\beta}^{(N-1)} = 0$ pour $\beta \neq \alpha$ et par suite $\lambda_{\alpha_1\sigma}^{(N-1)} = 1$, puisque $\sum_\beta \lambda_{\alpha_1\beta}^{(N-1)} = 1$.

Or,

$$\lambda_{\sigma_1\delta}^{(N)} = \sum_\gamma \lambda_{\alpha_1\gamma}^{(N-1)} \lambda_{\gamma\delta} = \lambda_{\alpha\delta}$$

On a donc

$$\lambda_{\alpha\delta} = \begin{cases} 1 & \text{si } \delta = \alpha_1, \\ 0 & \text{si } \delta \neq \sigma_1 \end{cases} \tag{71}$$

A tout indice α correspond un tel indice σ_1 ; c'est donc dire que tous les $\lambda_{\gamma\delta}$ sont égaux à 0 ou 1.

Sous-groupements cycliques. — Ceci étant, rangeons les groupements C_1, C_2, . . dans un nouvel ordre. A $C_1 = c_1$ correspond un et un seul groupement $C_{\delta_2} = c_2$, tel que $\lambda_{1\delta_2} = 1$. A C_{δ_2} correspond un groupement $C_{\delta_3} = c_3$ tel que $\lambda_{\delta_2\delta_3} = 1$, etc.

On extrait ainsi de $C_1, C_2, \ldots, C_\rho$, une suite de groupements $c_1, c_2, \ldots$, cette suite finit par épuiser toute la suite $C_1, \ldots, C_\rho$. Sans quoi on retrouverait dans la suite $c_1, c_2, \ldots$ une suite de moins de $\rho + 1$ termes $c_\mu, c_{\mu+1}, \ldots c_\nu$ dont les extrêmes coïncident : $c_\mu = c_\nu$. Or, d'après (71), chaque épreuve du troisième problème fera passer presque sûrement de l'un de ces c à l'un de ces c.

Alors si E_j appartient à c_μ, et si E_k appartient à un des groupements C_β autres que $c_\mu, c_{\mu+1}, \ldots, c_\nu$, on aura $\lambda_{\delta_\mu\beta}^{(n)} = 0$ pour tout n et, par suite, on aurait

$$Q_{jk}^{(n)} = \lambda_{\delta_\mu,\beta}^{(n)} P_k = 0 \qquad \text{et} \qquad \Pi_{jk} = \frac{Q_{jk}^{(1)} + Q_{jk}^{(2)} + \ldots + Q_{jk}^{(N)}}{N} = 0,$$

d'où $\Pi_{jk} = 0$, contrairement à l'hypothèse. Ainsi, en revenant aux notations primitives, on peut ranger $C_1, C_2, \ldots, C_\rho$ en un ordre tel que $\lambda_{l,l+1} = 1$ pour $l < \rho$ et $\lambda_{\rho 1} = 1$. Donc si E_j appartient à C_α et E_k à C_β, on aura

$$Q_{jk}^{(n)} = \begin{cases} P_k & \text{si } \beta - \alpha = n, \pmod \rho \text{ [1]}, \\ 0 & \text{si } \beta - \alpha \neq n, \pmod \rho. \end{cases} \tag{72}$$

---

[1] Comme d'habitude $a = b$, (mod ρ) signifie que $|a - b|$ est nul ou multiple entier de ρ.

Passons aux p_{ji}. On a, d'après (64),

$$Q_{jk}^{(2)} = \sum_i p_{ji} Q_{ik}^{(1)} = \sum_\gamma \sum_i^{C_\gamma} p_{ji} Q_{ik}^{(1)}$$

Si E_i, E_k appartiennent à C_γ, C_β, $Q_{ik}^{(1)}$ est nul pour $\gamma \not\equiv \beta - 1$, (mod ρ). En posant $\delta = \beta - 1$

$$Q_{jk}^{(2)} = \left[\sum_i^{C_\delta} p_{ji} \right] P_k$$

Or, $Q_{jk}^{(2)} = 0$ si $\beta - \alpha \not\equiv 2$ (mod ρ). Dans ce cas

$$0 = \sum_i^{C_\delta} p_{ji}$$

Donc $p_{ji} = 0$ quand E_j, E_i appartiennent à C_α, C_δ avec $\delta \not\equiv \alpha + 1$ (mod ρ).

Autrement dit, *en une épreuve du premier problème, le Système partant d'un état de* C_α *passe presque sûrement, à un état de* $C_{\alpha+1}$ (ou de C_ρ à C_1). De sorte que chaque épreuve opère presque sûrement une *permutation circulaire* sur les groupements C_1, C_2, ..., C_ρ. La part du hasard se restreint à la détermination de l'état d'arrivée parmi les états du groupement d'arrivée.

En raison de ce qui précède, nous appellerons $C_1, \ldots, C_\rho$ des groupements *cycliques* selon la dénomination proposée par M. Dœblin, pour une notion signalée d'abord en Algèbre par Frobenius et à laquelle, en Calcul des Probabilités, M. Hadamard [5], le premier, puis plusieurs auteurs, ont été indépendamment conduits.

D'ailleurs, par définition, $Q_{jk}^{(n)}$ est une fonction périodique de période N et, d'autre part, c'est, en vertu de (72), une fonction périodique de période ρ. Comme les P_k sont $\neq 0$, la succession des valeurs 0, 0, ..., 0, P_k, 0, 0, ..., 0, P_k, ..., de $Q_{jk}^{(n)}$ ne peut avoir une période plus petite que ρ. Si donc on a eu soin d'appeler N la plus petite des périodes de $Q_{jk}^{(n)}$, N sera nécessairement égal à ρ : *la période asymptotique* N (la plus petite) *des* $P_{jk}^{(n)}$ *est égale au nombre des groupements cycliques.*

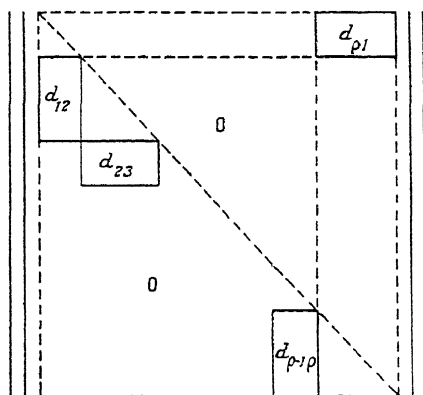
On voit aussi que l'on a pour les limites en moyenne

$$\Pi_k = \Pi_{jk} = \frac{Q_{jk}^{(1)} + \ldots + Q_{jk}^{(N)}}{N} = \frac{0 + \ldots + 0 + P_k}{N}$$

ou

$$\Pi_k = \frac{P_k}{N}. \tag{73}$$

Si l'on range les états en commençant par ceux de C_1, puis ceux de C_2, etc., la matrice des p_{jk} prendra une forme, où tous les

Fig. 4.

p_{jk} sont nuls en dehors des matrices d_{12}, d_{23}, ..., $d_{\rho-1\,\rho}$, $d_{\rho 1}$ ($d_{\alpha\beta}$ étant la matrice des p_{jk} où E_j appartient à C_α et E_k à C_β). Bien entendu la diagonale principale ne comportera que des zeros (toujours dans le cas où l'un des Q_{jk} est $=0$, c'est-à-dire où il y a plusieurs groupements cycliques).

Par exemple, Δ peut être

$$\begin{Vmatrix} 0 & 0 & 0 & 0 & 0 & p_{61} \\ 0 & 0 & 0 & 0 & 0 & p_{62} \\ p_{13} & p_{23} & 0 & 0 & 0 & 0 \\ p_{14} & p_{24} & 0 & 0 & 0 & 0 \\ p_{15} & p_{25} & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{36} & p_{46} & p_{56} & 0 \end{Vmatrix} \equiv \begin{Vmatrix} 0 & 0 & d_{31} \\ d_{12} & 0 & 0 \\ 0 & d_{23} & 0 \end{Vmatrix}$$

Si les p_{ij} non remplacés par zéro sont positifs, les sous-groupements cycliques sont, ici, dans leur ordre : (E_1, E_2), (E_3, E_4, E_5), (E_6).

Chaque épreuve faisant presque certainement passer d'un groupement C_α à celui qui le suit dans l'ordre circulaire, on voit qu'après

n épreuves, on aura passé presque certainement de C_α au groupement qui en diffère de n rangs dans l'ordre circulaire. Autrement dit

$$P_{jk}^{(n)} = 0 \quad \text{si} \quad \beta - \alpha \not\equiv n \quad (\text{mod } \rho).$$

Les sous-groupements cycliques dans le cas général. — Nous nous sommes placés pour définir les sous-groupements cycliques dans l'hypothèse où l'on se trouve dans le cas semi-régulier positif.

Si l'on passe maintenant au cas général, les états possibles se décomposent comme on a vu, en une suite de groupements (groupements de passage, groupements finals). A l'intérieur de chaque groupement final, on se trouve (*voir* p. 173) dans le cas semi-régulier positif, oscillant ou non. Par suite, *tout groupement final* g_ν, *dans lequel* $P_{jk}^{(n)}$ *ne converge pas pour chacun des couples d'états* E_j, E_k (*distincts ou non*) *appartenant à* g_ν, *se trouve décomposé en sous-groupements cycliques* $C_{\nu,1}, C_{\nu,2}, \ldots, C_{\nu,N_\nu}$, jouissant des mêmes propriétés que celles qui viennent d'être indiquées. En particulier, quand E_j, E_k appartiennent à un même groupement final g_ν, $P_{jk}^{(n)}$ est une fonction asymptotiquement périodique de n, dont la période asymptotique la plus petite, N_ν, est égale au nombre des sous-groupements cycliques constituant ce groupe final. La limite $Q_{jk}^{(n)}$ de $P_{jk}^{(n+tN_\nu)}$ quand t croît est zéro quand E_j, E_k appartiennent à deux sous-groupements cycliques $C_{\nu,\alpha}$. $C_{\nu,\beta}$ dont les rangs sont tels que $\beta - \alpha \not\equiv n$, (mod N_ν). Dans le cas contraire, elle est égale à $\Pi_k N_\nu$. On observe que chaque période N_ν est au plus égale au nombre des états du sous-groupement correspondant g_ν, donc $\leqq r$.

On voit que la matrice des p_{ik} peut maintenant se mettre, dans le cas le plus général, sous une forme analogue à la forme ci-contre (*fig*. 5), (où l'on a pris des valeurs déterminées qui pourraient être arbitraires, pour les indices d'états faisant partie de sous-groupements cycliques, de groupements de passages et de groupement finals). Les rectangles en traits pleins sont les matrices des sous-groupements cycliques. Les p_{ik} n'appartenant ni à ces matrices ni aux rectangles marqués d'un point d'interrogation (ici il n'y en a qu'un pour simplifier la figure), sont tous nuls.

Si, comme dans la figure 5, chaque groupement indécomposable comporte des sous-groupements cycliques, la diagonale principale de D sera nulle. Mais, bien entendu, il peut arriver que l'un ou l'autre

ou même aucun des groupements indécomposable ne comporte de sous-groupements cycliques. Il peut ne pas y avoir de groupement de passage. Il y a toujours au moins un groupement final, mais il peut

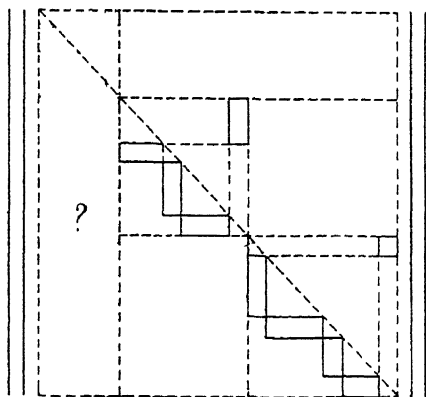



Fig. 5.

arriver que l'ensemble des états possibles forme lui-même un groupement final.

Critère du cas non-oscillant. — Pour qu'on soit dans le cas non-oscillant, il faut qu'aucun groupement final ne comporte de sous-groupement cyclique, et cela suffit.

En effet, si la condition n'était pas remplie, il existerait au moins un groupement final g_v dans lequel la plus petite période N_v des $P_{jk}^{(n)}$, étant égale (p. 184) au nombre des sous-groupements cycliques de g_v, serait supérieure à l'unité : on serait dans le cas oscillant.

Si, au contraire, la condition est remplie, $P_{jk}^{(n)}$ converge pour chaque couple d'états E_j, E_k. Quand E_k appartient à l'ensemble $\mathfrak{T}$ des groupements de passage, il sera même prouvé, page 192, que $P_{jk}^{(n)}$ converge vers zéro. Quand E_k appartient à un groupement final g_v, on a à distinguer deux cas.

Si E_j appartient aussi à un groupement final ou bien ce dernier est distinct de g_v et alors $P_{jk}^{(n)}$ converge encore vers zéro, ou bien E_j appartient aussi à g_v et, dire que g_v ne comporte pas de groupements cycliques, c'est dire que $P_{jk}^{(n)}$ est convergent.

Reste le cas où E_j appartient à $\mathfrak{T}$. Or, on a vu, page 180, que

$Q_{jk}^{(n+m)} = \sum_{i} Q_{ji}^{(n)} Q_{ik}^{(m)}$. On peut restreindre la sommation aux états E_i appartenant à g_{v} · car si E_i appartient à $\mathfrak{T}$, $Q_{ji}^{(n)} = 0$ d'après la page 194, et si E_i n'appartient ni à $\mathfrak{T}$, ni à g_{v}, $Q_{ik}^{(m)} = 0$.

Mais quand E_i appartient à g_v, $P_{ik}^{(t)}$ converge comme on vient de le rappeler, quand t croît de façon quelconque. C'est dire que $Q_{ik}^{(m)}$ est indépendant de m, dans les termes non nuls de la somme. Dès lors $Q_{jk}^{(n+m)}$ étant indépendant de m pour chaque valeur de n, $Q_{jk}^{(t)}$ est indépendant de t . autrement dit, $P_{jk}^{(n)}$ converge vers une seule limite quand n croît par valeurs entières quelconques.

Ceci nous donne une seconde démonstration du fait signalé plus haut, page 108; si la diagonale principale de D comporte moins de deux termes nuls, on se trouve nécessairement dans le cas non oscillant Car, dans le cas contraire, il y aurait au moins un groupement final comportant des sous-groupements cycliques, donc contenant plus d'un état; sa diagonale principale qui est nulle contiendrait au moins deux éléments. Nous voyons même maintenant que dans le cas oscillant toutes les diagonales de tous les groupements finals comportant des sous-groupements cycliques sont formées de zéros.

Remarque. — La répartition des états du Système, établie dans ce paragraphe sans invoquer une théorie plus générale, a été découverte par deux voies différentes Elle peut se déduire plus rapidement comme une application au Calcul des Probabilités de la théorie purement algébrique de Frobenius [1, 2, 3 et surtout 4] (1) concernant les matrices à termes ≥ 0. C'est la méthode qui a été suivie en 1931 par M. de Mises [1, p. 537-549]. A ses résultats sont venus s'ajouter ceux qui ont été obtenus en 1935, au moyen de la même méthode par M. Romanovski [3] ·

Toutefois, dans le cas du battage des cartes, c'est M. Hadamard [5] qui a découvert la répartition en sous-groupements cycliques, en 1928, grâce à une méthode nouvelle, employée ensuite par d'autres auteurs, et que nous allons exposer maintenant.

---

(1) Ou bien voir un résumé par M. Romanovski [3, p. 147-179] des résultats de Frobenius.

TROISIÈME MÉTHODE.

MÉTHODE DIRECTE.

Introduction. — M. Hadamard [5] est le premier à avoir employé pour le problème des probabilités en chaîne, une méthode directe permettant d'utiliser exclusivement le langage des probabilités et d'éviter tout emprunt aux théories algébriques. Sa méthode, qu'il avait appliquée au battage des cartes, s'étend sans modification au cas du problème de Markoff où l'on a la condition $(T'_1) = \sum p_{il} = 1$

Elle lui a permis d'établir des résultats non encore obtenus à cette époque par la méthode algébrique et qui mettent en évidence la distinction entre les particularités certaines (ou de probabilité égale à 1) des variations du système considéré et celles qui sont strictement dues au hasard.

M. Kolmogoroff, s'inspirant de la méthode de M. Hadamard, a pu la modifier de façon à l'étendre au cas où la condition (T'_1) n'est pas réalisée, ce qui nécessite des modifications importantes et non évidentes du raisonnement de M. Hadamard. Par la même méthode, il a pu même retrouver plusieurs résultats déduits par M. de Mises [1] et Romanovsky [3] des résultats algébriques de Frobenius. Après avoir exposé, il y a longtemps, ces divers points dans son cours, il en a publié récemment un résumé, sans démonstration, puis un exposé complet (Kolmogoroff, 3, 4) (1).

Quelque temps avant la publication de ce résumé, M. Dœblin [3] m'avait communiqué une méthode et des résultats très analogues obtenus dans l'ignorance des travaux de M. Hadamard déjà publiés et, bien entendu, de ceux encore inédits de M. Kolmogoroff (2). Il y a récemment ajouté un grand nombre de résultats nouveaux.

Plusieurs des résultats de MM. Kolmogoroff et Dœblin avaient été obtenus par M. Romanovsky [3] en poussant plus loin la méthode algébrique exposée ci-dessus.

---

(1) M. Kolmogoroff a pu obtenir des démonstrations applicables au cas où l'ensemble des états possibles est dénombrable (fini ou non).

(2) *Voir* aussi A. Fouillade [1, 2].

L'avantage de la méthode directe, c'est qu'elle fait mieux comprendre ce qui se passe, même pour un petit nombre d'épreuves. Il faut cependant reconnaître les avantages de la méthode algébrique qui sont : de synthétiser comme on a pu le faire plus haut des détails fournis par la méthode directe et qui ne sont pas essentiels, de fournir une expression explicite des probabilités itérées, enfin de se prêter, comme on le verra dans les notes A, B, C, D à la fin de l'ouvrage ou comme le montrent les travaux de Frobenius, à des extensions utiles à des théories distinctes du Calcul des Probabilités.

La méthode directe permet d'établir un grand nombre des résultats qui ont été prouvés par d'autres voies dans les pages précédentes. Il ne nous a pas paru nécessaire de reproduire ici toutes ces démonstrations nouvelles qu'on pourra trouver dans les mémoires cités plus haut.

États conséquents. — Il sera commode, dans la suite, d'emprunter une expression introduite par Poincaré dans la théorie des trajectoires et reprise ici, en raison de l'analogie des circonstances, par M. Hadamard. Appelons *conséquent* d'ordre n d'un état E_j un état E_k tel que $P_{jk}^{(n)} \neq 0$. Comme

$$P_{jl}^{(n+m)} = \sum_h P_{jh}^{(n)} P_{hl}^{(m)},$$

il est clair que si E_h est conséquent d'ordre n de E_j et E_l conséquent d'ordre m de E_h, alors E_l est consequent d'ordre $n+m$ de E_j. Nous dirons en outre qu'un état E_l est conséquent de E_j s'il existe un entier s tel que $P_{jk}^{(s)} \neq 0$, c'est-à-dire si E_l est conséquent de E_j d'au moins un ordre s (quelle que soit cette valeur de s).

On peut transformer ces définitions. On a $P_{jk}^{(2)} = \sum_{i=1}^{i=r} p_{ji} p_{ik}$ et l'on voit facilement (par récurrence ou en vertu du théorème des probabilités totales) que

$$P_{jk}^{(n)} = \sum_{i_1=1}^{i_1=r} \cdots \sum_{i_{n-1}=1}^{i_{n-1}=r} p_{ji_1} p_{i_1 i_2} \cdots p_{i_{n-1}k}.$$

Si donc $P_{jk}^{(n)} \neq 0$, il existe au moins une suite de $n-1$ états (distincts

ou non) $E_{i_1}, \ldots, E_{i_{n-1}}$ tels que le produit $p_{ji_1}p_{i_1i_2}\cdots p_{i_{n-1}k} \neq 0$. Réciproquement s'il existe au moins une telle suite, $P^{(n)}_{jk}$ est $\neq 0$.

On peut aussi observer que le produit $p_{ji_1}p_{i_1i_2}\cdots p_{i_{n-1}k}$ est la probabilité pour que le système passe de E_j à E_k précisément par E_{i_1} à la première épreuve, par E_{i_2} à la seconde ... par E_k à la $n^{\text{ième}}$. Nous dirons avec M. Dœblin qu'il y a un chemin $E_j, E_{i_1}, \ldots, E_{i_{n-1}}, E_k$ de E_j à E_k si cette probabilité est $\neq 0$, et nous dirons que ce chemin est d'ordre n. On peut toujours supposer qu'un chemin est d'ordre au plus égal au nombre r des états possibles. Il suffit évidemment de prouver que si un chemin existe de E_j à E_k, il en existe un autre entre E_j et E_k dont les éléments précédant E_k sont distincts (E_j, E_k étant distincts ou non). Or, si le premier chemin contenait deux éléments identiques $E_{i_k} = E_{i_{k+l}}$, alors le produit

$$p_{ji_1}\cdots p_{i_{k-1}i_k}p_{i_ki_{k+1}}\cdots p_{i_{k+l-1}i_k}p_{i_ki_{k+l+1}}\cdots p_{i_{n-1}k}$$

étant $\neq 0$, il en serait de même de

$$p_{ji_1}\cdots p_{i_{k-1}i_k}p_{i_ki_{k+l+1}}\cdots p_{i_{n-1}k}$$

et l'on aurait supprimé de la suite des indices au moins i_{k+l}. En opérant ainsi tant qu'il y aura dans la suite des états identiques, on arrivera bien à un chemin composé d'états différents et par suite de r états au plus précédant E_k.

Condition pour que l'ensemble, G, des états possibles soit indécomposable. — On a vu page 169, que, si G est indécomposable, les Π_{jk} sont tous $\neq 0$. Par suite, pour chaque couple ordonné j, k, il existe au moins un entier n tel que $P^{(n)}_{jk}$ soit $\neq 0$, c'est-à-dire que E_k est conséquent de E_j.

Réciproquement; si tout état est conséquent de tout autre, distinct ou non du premier, il est bien clair que G est indécomposable puisque dans le cas contraire, il y aurait, comme on l'a vu plus haut, au moins deux indices k', k'' tels que $P^{(n)}_{k''k'} = 0$, quel que soit n, tels par suite que k'' ne soit pas conséquent de k'.

Ainsi, comme l'a démontré M. de Mises [1], qui l'a déduit d'un théorème algébrique de Frobenius, *pour que* G *soit indécomposable il faut et il suffit que quels que soient les états* E_h *et* E_k, *distincts ou non, chacun d'eux soit conséquent de l'autre*, et nous savons

même qu'ils seront alors conséquents d'ordre $\leqq r$. (Ce résultat reste valable si l'on remplace G par un groupement final.)

On peut donc dire aussi avec M. Dœblin qu'un tableau $\|p_{ik}\|$ est indécomposable si, quels que soient les indices j et k, il existe une suite d'un nombre convenable (qu'on peut supposer variable avec j et k, mais $\leqq r$) d'indices $i_1, i_2, \ldots, i_{n-1}$ tels que le produit $p_{ji_1}p_{i_1i_2}\ldots p_{i_{n-1}k}$ soit $\neq 0$. C'est là une définition purement algébrique et qui, précisément pour cela, peut être utile pour reconnaître les tableaux indécomposables.

On peut généraliser : nous considérerons G comme décomposable en ν épreuves, s'il est possible de décomposer G en deux groupements A′, B′ comprenant chacun au moins un élément et tels que pour tout couple d'indices, l'un k' de A′, l'autre k'' de B′, on ait $P_{k''k'}^{(\nu)} = 0$. En raisonnant alors sur les $P_{ij}^{(n\nu)}$ comme on vient de le faire sur les $P_{ij}^{(n)}$, on arrive au résultat suivant :

Pour que G soit indécomposable en ν épreuves, il faut et il suffit que pour tout couple d'états E_h, E_k (distincts ou non), chacun d'eux soit conséquent de l'autre, à un ordre égal à ν ou multiple de ν, ou encore il faut et il suffit d'après la page 169, qu'il n'existe aucun couple E_j, E_l d'états distincts ou non tels que

$$\lim_{n \to \infty} P_{jl}^{(n\nu)} = 0$$

Propriété des groupements indécomposables. — On a vu que, si G est indécomposable, — et d'après la page 177, qu'au cas contraire, si l'on ne s'adresse qu'à un groupement final, — deux états quelconques sont chacun conséquent de l'autre. Cette propriété appartient aussi aux groupements de passage comprenant plus d'un état.

En effet, supposons qu'un même groupement indécomposable g comprenne deux états distincts dont l'un E_j n'est pas conséquent de l'autre E_h. Soient alors b l'ensemble de E_h et des états de g qui sont conséquents de E_h, a l'ensemble des autres états de g. a et b comprennent chacun un élément, au moins, et l'on a $p_{k''k'} = 0$ quand k'' appartient à b et k' à a. (Car dans le cas contraire k' serait conséquent de k'', donc de E_h contrairement à la définition de k'.) Dès lors g serait décomposable contrairement à l'hypothèse. Si g comprend plus d'un élément, deux éléments distincts E_i, E_l étant conséquents

l'un de l'autre, chacun d'eux sera conséquent de lui-même au moins par l'intermédiaire de l'autre. Ainsi, quand g comprend plus d'un élément, deux quelconques de ses éléments, distincts ou non, sont conséquents l'un de l'autre.

Construction directe des groupements indécomposables. — Soit E_i un état quelconque; dans la décomposition opérée plus haut, il appartient à un groupement indécomposable g_m bien déterminé. On peut déterminer directement g_m sans connaître la décomposition totale de G en $g_1, g_2, \ldots, g_l$.

En effet, soit γ_i l'ensemble formé de E_i et des états E_k de G, s'il en existe, qui sont conséquents de E_i et dont E_i est conséquent. Les états E_k étant conséquents de E_i, ils appartiennent à g_m, à g_{m+1}, ou . ou g_l. E_i étant leur conséquent, ils appartiennent donc a g_1, ou g_2, . , ou g_m. Ils appartiennent donc à g_m Ainsi γ_i appartient à g_m.

Mais nous avons vu que si g_m comprend plus d'un élément, tous ses états sont conséquents l'un de l'autre Dès lors γ_i comprend g_m γ_i est donc identique à g_m. Ainsi la décomposition en groupements indecomposables peut s'obtenir directement de la façon suivante (ce qui montrera en même temps qu'il n'existe qu'une telle décomposition de G).

A tout état E_i de G, on associe un groupement γ_i défini comme plus haut Chaque γ_i est un des groupements indécomposables $g_1, g_2, \ldots, g_l$, construit indépendamment des autres. Dès lors, si γ_i et γ_k correspondent à deux états E_i, E_k : ou bien γ_i et γ_k sont disjoints, c'est-à-dire sans etat commun, ou bien ils sont identiques. On obtient la décomposition de G en groupements indécomposables en formant tous les γ_i et en ne retenant que ceux qui sont distincts. Au lieu de former tous les γ_i, on pourra former γ_1; puis choisir un élément E_{i_2} de G, n'appartenant pas à γ_1, s'il en existe, et former γ_{i_2}; puis choisir E_{i_3}, s'il en existe n'appartenant ni à γ_1 ni à γ_{i_2} et former γ_{i_3}, etc. La suite γ_1, $\gamma_{i_2}, \gamma_{i_3}, \ldots$ sera celle des groupements indécomposables dans un ordre quelconque.

C'est la construction indiquée par MM. Kolmogoroff [3, 4] et Dœblin [3].

Probabilité de sortir des groupements de passage. — Supposons

que E_i appartienne à l'ensemble $\mathfrak{T}$ des états des groupements de passage Il appartient, par exemple, à g_α ($\alpha \leqq u$); or, il existe un état E_j, au moins, de g_α, tel que $\sum_k^{g_\alpha} p_{jk} < 1$ et, par suite, il existe un état E_{k_0} n'appartenant pas à g_α et tel que $p_{jk_0} \neq 0$ D'ailleurs, on a vu (p. 174) que p_{jk_0} serait nul si E_{k_0} appartenait à g_1 ou g_2, ..., ou $g_{\alpha-1}$ Donc E_{k_0} appartient à $g_{\alpha+1}$, ou $g_{\alpha+2}$, .., ou g_l. D'autre part, E_j est conséquent de E_i, dont E_{k_0} aussi : tout état E_i de $\mathfrak{T}$ a au moins un conséquent E_{k_0} appartenant à un groupement indécomposable de rang supérieur à celui auquel appartient E_i Ou bien E_{k_0} n'appartient pas à $\mathfrak{T}$, ou bien on peut opérer sur E_{k_0} comme sur E_i et trouver un conséquent de E_i appartenant à un des groupements indécomposables $g_{\alpha+2}$, $g_{\alpha+3}$, ..., et ainsi de suite. Finalement E_i a un conséquent E_k n'appartenant pas à $\mathfrak{T}$. Il existe un rang $\nu \leqq r$, tel que $P_{ik}^{(\nu)} \neq 0$ Alors si $c_i^{(r)}$ est la probabilité que le Système partant de E_i se trouve à la $r^{\text{ième}}$ épreuve dans le groupement final g_ρ dont fait partie E_k, on aura $c_i^{(r)} \neq 0$. Car, ou bien $\nu = r$ et $c_i^{(r)} \geqq P_{ik}^{(r)} > 0$; ou bien $\nu < r$ et alors $c_i^{(r)}$ est au moins égal à la probabilité que le système partant de E_i arrive à E_k après ν épreuves et que partant de l'état E_k de g_ρ, il soit encore après $r - \nu$ épreuves sur g_ρ. Ce dernier événement étant presque certain, on aura $c_i^{(r)} \geqq P_{ik}^{(\nu)} > 0$.

Mais la probabilité $a_i^{(r)}$ que le système partant de E_i soit hors de $\mathfrak{T}$ à la $r^{\text{ième}}$ épreuve est $\geqq c_i^{(r)}$, elle est donc positive, et si a est le plus petit des nombres $a_i^{(r)}$ quand E_i parcourt $\mathfrak{T}$, on a $a_i^{(r)} \geqq a > 0$. Soit maintenant $b_i^{(n)} = 1 - a_i^{(n)}$ la probabilité que le système partant de E_i de $\mathfrak{T}$ appartienne encore à $\mathfrak{T}$ après $n > t\,r$ épreuves On a $b_i^{(r)} \leqq 1 - a$. On aura même $b_i^{(n)} \leqq (1 - a)^t$.

Pour le montrer nous ferons usage d'une observation évidente qui nous sera souvent utile.

Digression. — Si un événement E consiste dans le concours de deux événements A et B, si A consiste dans la réalisation de l'un quelconque des événements incompatibles A_1, . , A_s, enfin, si la probabilité de B quand A_i a lieu est comprise quel que soit i entre m et M, alors on a

$$(\text{Prob. A})\,m \leqq \text{Prob. E} \leqq (\text{Prob. A})\,\text{M},$$

comme il résulte de l'expression de Prob. E par une somme de probabilités composées.

Nous savons que la probabilité pour que le Système revienne en une epreuve d'un état d'un groupe final à un état n'appartenant pas à ce groupe final est nulle; il en est donc de même de la probabilité que le systeme partant de l'ensemble des groupes finals passe à $\mathfrak{T}$ en un nombre quelconque d'épreuves. Dès lors $b_i^{(n)}$ est donc égal à la probabilité pour que le Système partant de E_i de $\mathfrak{T}$ soit encore sur $\mathfrak{T}$ après $n-r$ épreuves et soit encore sur $\mathfrak{T}$ après r autres épreuves. La probabilité que ce dernier événement ait lieu en partant d'un état E_h de $\mathfrak{T}$ est $\leqq 1-\alpha$. En vertu de l'observation générale faite ci-dessus, on aura donc, en faisant varier E_h sur $\mathfrak{T}$,

$$b_i^{(n)} \leqq b_i^{(n-r)}(1-\alpha) \qquad \text{d'où} \quad b_i^{(n)} \leqq b_i^{(n-tr)}(1-\alpha)^t \leqq (1-\alpha)^t.$$

Dès lors

$$\lim_{n \to \infty} b_i^{(n)} = 0 \qquad \text{d'où} \quad \lim_{n \to \infty} a_i^{(n)} = 1$$

Or, $a_i^{(n)}$ est une fonction non décroissante de n. Car $a_i^{(n-1)}$ est au moins égal à la probabilité pour que le Système partant de l'état E_i de $\mathfrak{T}$, sorte de $\mathfrak{T}$ après n épreuves et qu'il n'y rentre pas à la $(n+1)^{\text{ième}}$ Donc $a_i^{(n+1)} \geqq a_i^{(n)} \times 1$.

Dés lors la probabilité que le Système partant d'un état E_i de $\mathfrak{T}$ en sorte au bout d'un nombre (aléatoire) fini d'épreuves est la limite de $a_i^{(n)}$, elle est donc égale à 1.

D'ailleurs, si E_i n'appartient pas à $\mathfrak{T}$, E_i appartient à un groupement final, il y a donc une probabilité nulle que le Système partant de E_i se trouve dans $\mathfrak{T}$ après une épreuve et de même après un nombre quelconque d'épreuves. Dans ce cas, $b_i^{(n)} = 0$ quel que soit n. Ainsi se trouve établi ce résultat de M. Romanovsky [3], obtenu, auparavant, dans un cas particulier par M. de Mises [1], retrouvé ensuite autrement et indépendamment par MM. Kolmogoroff et Dœblin : *la probabilité que le Système partant d'un état possible quelconque* E_j *se trouve, après n épreuves, appartenir à l'ensemble des groupements de passage, tend vers zéro avec* $\frac{1}{n}$. Ceci justifie bien la dénomination des groupements de passage.

En particulier, il en résulte que

$$\lim_{n \to \infty} P_{jh}^{(n)} = 0,$$

quel que soit j, quand E_h appartient à un groupement de passage.

On a, d'ailleurs, comme on l'a vu page 176, $P_{jh}^{(n)} = 0$ quel que soit n quand, E_j appartenant à un groupement final, E_h n'appartient pas à celui-ci. Enfin, il résulte de ce qu'on a vu plus haut, page 177, que les limites en moyenne Π_{jh} sont $\neq 0$ et indépendantes de j quand h et j appartiennent à un même groupement final. On aura donc pour la matrice des Π_{jh} la disposition ci-dessous, (*fig.* 6), où les matrices $A'_{u+1}, \ldots, A'_t$ correspondent aux groupements finals $g_{u+1}, \ldots, g_t$.

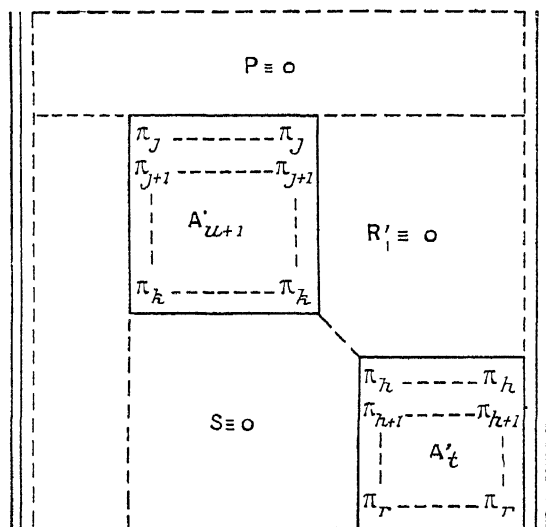



Fig. 6.

Cas semi-régulier. — On voit que pour qu'on soit dans le cas semi-régulier, il faut d'abord que les régions R'_1 et S, de la figure 6, disparaissent, donc qu'il n'y ait qu'un seul groupement final. Soient $E_j, E_{j+1}, \ldots E_r$ ses états.

On a vu que

$$\Pi_{kl} = \sum_{i=1}^{i=r} \Pi_{ki} \Pi_{il}, \qquad \sum_{i=1}^{i=r} \Pi_{ki} = 1.$$

Or, $\Pi_{ki} = 0$ pour $i < j$ et $\Pi_{il} = \Pi_l$ pour $i \geqq j$, $l \geqq j$. On a donc, pour $l \geqq j$,

$$\Pi_{kl} = \left[\sum_{i=j}^{i=r} \Pi_{ki}\right] \Pi_l \qquad \text{et} \qquad \sum_{i=j}^{i=r} \Pi_{ki} = 1,$$

d'où

$$\Pi_{kl} = \Pi_l \qquad \text{pour} \quad l \geqq j.$$

Cette égalité est donc vraie quel que soit k. Comme on a aussi $\Pi_{kl} = 0$, quel que soit k pour $l < j$, on voit bien que la condition est suffisante

Pour qu'en outre, les Π_k soient tous $\neq 0$, c'est-à-dire pour qu'on soit dans le cas semi-régulier positif, il faut évidemment que, de plus, les groupements de passage (pour lesquels les Π_{ik} sont nuls) disparaissent. Dès lors, il faut que l'ensemble des états possibles soit indécomposable.

Cette condition est évidemment suffisante.

Démonstration directe de la répartition d'un groupement indécomposable en sous-groupements cycliques — Faisons une remarque préliminaire. Considérons un état E_i d'un groupement indécomposable g comprenant plus d'un élément. Il est à lui-même son conséquent, de divers ordres $m_1 < m_2 < \ldots$. Soit d_i le plus grand commun diviseur de $m_1, m_2, \ldots$ (comme on sait que $m_1 \leqq r$, on a $1 \leqq d_i \leqq r$) Soit maintenant d_k le nombre correspondant de la même façon à un état E_k de g. On sait que E_k est conséquent d'un certain ordre, n' de E_i et E_i, d'un certain ordre n'' de E_k. Donc E_i est conséquent de lui-même d'ordre $n = n' + n''$ et, par suite, n est divisible par d_i De même, il est divisible par d_k.

Mais on peut dire aussi : E_i est conséquent d'ordre n'' de E_k, E_i est conséquent d'ordre m_t de lui-même, E_k est conséquent d'ordre n' de E_i, donc E_k est conséquent de lui-même d'ordre $n + m_t$ et par suite $n + m_t$ et (comme nous l'avons vu) n sont divisibles par d_k. Dès lors m_t est divisible par d_k; ceci ayant lieu pour $m_1, m_2, \ldots, d_i$ est divisible par d_k et de même d_k par d_i. On a donc finalement $d_i = d_k$.

Les nombres d_i attachés aux états E_i d'un même groupement indécomposable ont une valeur commune; appelons N_ρ la valeur commune ainsi attachée aux états du $\rho^{\text{ième}}$ groupement indécomposable g_ρ.

Soient maintenant deux états E_i, E_k de g_ρ. E_k est conséquent de E_i;

s'il peut être considéré comme conséquent de E_l d'ordre m' et m'', comme, inversement, E_l est conséquent de E_k d'au moins un ordre n'', E_k est conséquent de lui-même, à la fois des ordres $n''+m'$ et $n''+m''$, ces deux ordres sont divisibles par N_v, donc $m'-m''$ est divisible par N_v, ce qu'on peut écrire

$$m'' \equiv m', \quad (\text{mod } N_v),$$

et qu'on exprime en disant : m' et m'' sont congrus (mod N_v) Il y a donc un entier α pour lequel $1 \leqq \alpha \leqq N_v$ tel que

$$m' \equiv \alpha, \quad (\text{mod } N_v), \qquad m'' \equiv \alpha, \quad (\text{mod } N_v)$$

Prenons alors pour E_l un état arbitraire mais fixe, E_{l_0}, de g_v et appelons $C_{v,\alpha}$ l'ensemble des états E_k de g_v tels que les ordres des chemins de E_{l_0} à E_k — qui sont tous congrus entre eux, module N_v, — soient congrus à α. On voit que g_v se trouve décomposé en un nombre fini d'ensembles disjoints d'états, les ensembles $C_{v,1}$, $C_{v,2}$, ..., C_{v,N_v}. On peut aussi poser $C_{v,n} = C_{v,\alpha}$ quand n est congru à α, (mod. N_v).

Considérons d'abord le cas où le groupement g_v est un groupement final. Les conséquents d'ordre 1 de $C_{v,n}$ sont conséquents d'ordre $n+1$ de E_{l_0}, donc ils appartiennent tous à $C_{v,n+1}$.

De sorte que toute épreuve fait passer presque certainement tout état de $C_{v,n}$ sur $C_{v,n+1}$, à savoir $C_{v,1}$ sur $C_{v,2}$; $C_{v,2}$ sur $C_{v,3}$ ... , C_{v,N_v-1} sur C_{v,N_v}, puis C_{v,N_v} sur $C_{v,1}$, $C_{v,1}$ sur $C_{v,2}$, et ainsi de suite, indéfiniment. Il en résulte en particulier que chacun des $C_{v,\alpha}$ existe, c'est-à-dire comprend au moins un élément.

C'est ce qu'on exprime en disant que les sous-groupements $C_{v,n}$ sont des *sous-groupements cycliques* de g_v. On voit que l'effet d'une épreuve sur un état appartenant au groupe final g_v est la résultante de deux actions : une action presque certaine faisant passer le Système d'un sous-groupement cyclique au suivant dans leur ordre circulaire, une action aléatoire qui détermine sa position fortuite dans ce nouveau sous-groupement certain.

Dans le cas où g_v est un groupement de passage, les choses sont un peu moins simples. Les conséquents d'ordre 1 de $C_{v,n}$, n'appartenant pas nécessairement tous à g_v, sont répartis, non seulement, dans $C_{v,n+1}$, mais aussi à l'extérieur de g_v dans g_{v+1}, ..., g_t.

D'ailleurs les sous-groupements cycliques d'un groupement de

passage auraient beaucoup moins d'intérêt que ceux d'un groupement final, puisque d'après la page 194, la périodicité disparaît à la limite pour les premiers alors qu'elle se conserve pour les seconds.

Critères des différents cas. — Nous avons vu, pages 194, 173 et 185, que : *pour qu'on soit dans le cas semi-régulier, il faut et il suffit qu'il n'existe qu'un seul groupement final,*

Pour qu'on soit dans le cas semi-régulier positif, il faut et il suffit que l'ensemble des états possibles soit indécomposable,

Pour qu'on soit dans le cas non oscillant, il faut et il suffit qu'il n'y ait de sous-groupements cycliques dans aucun groupement final.

Dès lors :

Pour qu'on soit dans le cas régulier, il faut et il suffit qu'il n'existe qu'un groupement final et que, dans celui-ci, il n'existe aucun sous-groupement cyclique,

Pour qu'on soit dans le cas positivement régulier, il faut et il suffit que l'ensemble des états possibles soit indécomposable et ne comporte aucun groupement cyclique.

Application des critères. — M. Dœblin [3] a fait une observation pouvant faciliter l'application des critères précédents et des critères d'autres formes donnés plus haut.

Une modification quelconque des valeurs des p_{ik} qui respecte les conditions

$$p_{ik} \geqq 0, \qquad \sum_k p_{ik} = 1,$$

qui remplace des valeurs $\neq 0$ par des valeurs $\neq 0$ et des valeurs nulles par des valeurs nulles, ne change évidemment rien aux relations de « conséquence » et par suite ne modifie ni les groupements indécomposables, ni les sous-groupements cycliques. Dès lors, on pourra simplifier l'application des critères ci-dessus en remplaçant dans la matrice des p_{ik} tout $p_{ik} \neq 0$, par le signe $+$. Une étude systématique de la matrice ainsi simplifiée permettra d'abord de discerner les conséquents d'ordre $\leqq r$ de chaque état et ensuite d'en profiter pour déterminer comme à la page 191 les groupements indécompo-

sables g_v. On distinguera facilement parmi ceux-ci les groupements finals; savoir tout g_v tel que tous ses conséquents lui appartiennent. Puis, dans chaque groupement final, on pourra voir s'il y a plusieurs sous-groupements cycliques. Il faut d'abord pour cela que la diagonale principale soit toute nulle, ce que l'on voit de suite. On déterminera ensuite les sous-groupements cycliques en appliquant le procédé de la page 191.

On voit que l'avantage de ces critères et de cette méthode, c'est qu'elle ne nécessite que de la patience, mais aucun calcul. Toutefois observons que pour la détermination des cas semi-réguliers, le critère algébrique indiqué page 111 présente ce grand avantage que les calculs — qui se trouvent réduits à la résolution d'équations linéaires, — fournissent du même coup les valeurs des probabilités limites en moyenne Π_k, valeurs dont on aura souvent besoin.

Principes ergodique, presque ergodique. — On dit généralement que le principe ergodique est satisfait quand les $P_{jk}^{(n)}$ convergent vers des limites P_k indépendantes de l'état initial E_j; c'est-à-dire dans le cas régulier. Dans le cas examiné jusqu'ici, où il y a un nombre fini d'états et une suite discrète d'épreuves, il est équivalent de dire avec avec M. Kolmogoroff que le principe ergodique est vérifié si $P_{ik}^{(n)} - P_{jk}^{(n)}$ converge vers zéro avec $\frac{1}{n}$ quels que soient i, j, k. Cette condition est évidemment vérifiée quand on est dans le cas régulier. Réciproquement, si elle est vérifiée, alors il est clair que la limite en moyenne $\Pi_{ik} - \Pi_{jk}$ de $P_{ik}^{(n)} - P_{jk}^{(n)}$ doit être nulle. Donc on doit se trouver dans le cas semi-régulier. Si, d'autre part, la période asymptotique des $P_{ik}^{(n)}$ n'était pas égale à 1, il y aurait des sous-groupements cycliques d'un même groupement final. En prenant i, j, k dans ce groupement final, on aurait, quand n croît par valeurs s différant entre elles de multiples de N, une limite égale à $0 - P_k$ ou à $P_k - 0$ quand E_j appartient au même groupement cyclique que E_k et non E_i, ou inversement. Or, $P_k \neq 0$, d'où la contradiction annoncée.

M. Dœblin [3] a observé que, dans le cas où l'ensemble des états possibles est indécomposable, se trouve vérifié un principe « presque ergodique ». On voit d'abord, en effet, que s'il n'y a pas de sous-groupements cycliques, on a même le principe ergodique. Quand, au contraire, il y en a, alors, on a vu que $P_{ik}^{(n)} - P_{jk}^{(n)}$ ne tend pas vers

zéro quand i, j, k sont quelconques. Mais si E_i et E_j appartiennent à un même groupement cyclique, $P_{ik}^{(n+t\lambda)} - P_{jk}^{(n-t\lambda)}$ tend, quand t croît, vers $Q_{ik}^{(n)} - Q_{jk}^{(n)}$ qui (étant égal à $P_k - P_k$ quand n a certaines valeurs et à o — o pour les autres) est nul.

On peut donc formuler un principe « presque ergodique » en disant : $P_{ik}^{(n)} - P_{jk}^{(n)}$ tend vers zéro quand E_i et E_j appartiennent à un même sous-groupement cyclique, et ceci quel que soit E_k. C'est en particulier ce qui a lieu si E_i, E_j sont conséquents du même ordre d'un même état de E_l On peut donc dire aussi : dans le cas où l'ensemble des états possibles est indécomposable, si l'on connaît l'état initial E_l du système et si l'on appelle E_i, E_k les états du Système après m et $n+m$ épreuves, la connaissance exacte de la position intermédiaire E_i n'influe pas sur le comportement asymptotique de la probabilité $P_{ik}^{(n)}$.

Les mêmes remarques se présentent dans le cas général quand on ne considère que ce qui se passe à l'intérieur d'un groupement final

Cas du battage des cartes. — Dans le cas où se trouve réalisée la condition

$$\text{(T)} \qquad \sum_j p_{jk} = 1.$$

(et en particulier dans le cas du battage des cartes), certaines simplifications se produisent.

Tout d'abord, il n'y a plus de groupements de passage, car on a vu que $P_{ik}^{(n)}$ tend vers zéro si E_k appartient à un tel groupement; or, ceci ne peut avoir lieu quel que soit i en raison de l'égalité

$$\sum_i P_{ik}^{(n)} = 1$$

D'autre part, si a_ν est le nombre des états d'un groupement final g_ν auquel appartient E_k, on a

$$1 = \sum_{j=1}^{j=r} \Pi_{jk} = a_\nu \Pi_k, \qquad \text{d'où} \qquad \Pi_k = \frac{1}{a_\nu},$$

de sorte que la limite en moyenne de $P_{jk}^{(n)}$, qui est indépendante de la position de l'état initial E_j dans son groupement final g_ν, est aussi

indépendante de la position, dans le même groupement final g_v, de l'état final E_k.

Enfin, le nombre $m_{v,\alpha}$ des états d'un sous-groupement cyclique $C_{v,\alpha}$ d'un même groupement final g_v est indépendant de α. Car on a vu [formule (73), page 183] que si E_k appartient à $C_{v,\alpha}$

$$Q_{kk}^{(\lambda)} = P_k = N_v \Pi_k = \frac{N_v}{a_v},$$

et l'on a

$$1 = \sum_{j=1}^{j=r} Q_{jk}^{(\lambda)} = \sum_{j}^{C_{v,\alpha}} Q_{jk}^{(\lambda)} = \sum_{j}^{C_{v,\alpha}} P_k = m_{v,\alpha} \frac{N_v}{a_v}.$$

Donc $m_{v,\alpha}$ est indépendant de α, $m_{v,\alpha} = m_v$. On a, de plus, bien entendu,

$$a_v = m_v N_v \tag{74}$$

Nous avons vu qu'on peut ranger les états de telle sorte que la matrice des p_{ik} prenne la forme de la figure 5, page 185. Mais dans le cas du battage des cartes, on devra y supprimer la partie du tableau (à gauche comme en haut) qui correspondait à l'existence éventuelle, devenue ici impossible, des groupements de passage. Il faudra aussi remplacer les matrices rectangulaires correspondant aux sous-groupements cycliques par des matrices carrées; tous les p_{ik} situés en dehors de celles-ci sont nuls.

Stabilité à la Poisson. — MM. Kolmogoroff [3, 4] et Dœblin [3] ont démontré la propriété suivante des groupements finals, propriété qu'on peut rapprocher de la stabilité « à la Poisson » en Mécanique .

Soient E_j, E_k *deux états quelconques d'un même groupement final; quand le Système part de* E_j, *il passe presque sûrement une infinité de fois par* E_k (*que* E_j, E_k *soient ou non distincts*).

La probabilité Q pour que E_k ne soit obtenu qu'un nombre (aléatoire) fini de fois à partir de E_j est la somme

$$q_{j0} + q_{j1} + \ldots + q_{jn} + \ldots$$

de la probabilité q_{j0} que E_k ne soit jamais obtenu à partir de E_j et des probabilités q_{jn} que E_k soit obtenu à la $n^{\text{ième}}$ épreuve à partir de E_j et jamais après. Or, $q_{jn} = P_{jk}^{(n)} q_{k0}$.

Il suffit donc de démontrer que q_{j0} et q_{k0} sont nuls, ou encore que $q_{l0} = 0$, quel que soit l'état E_l du même groupement final g_α, que celui contenant E_j, E_k.

A cet effet, observons que les deux états E_l, E_k de g_α sont conséquents l'un de l'autre, d'un ordre $\nu \leq r$

La probabilité $w_{lk}^{(n)}$ que le Système partant de E_l passe par E_k au moins une fois au cours de n épreuves est évidemment une fonction non décroissante de n, positive pour $n = \nu$, donc positive pour $n = r$. Dès lors les nombres $w_{lk}^{(r)}$ étant tous $\neq 0$, le plus petit de ces nombres, quand E_l, E_k parcourent g_α, est un nombre positif w.

Ceci étant, si $n > tr$, on aura

$$1 - w_{lk}^{(n)} \leq (1 - w)^t$$

Car $1 - w_{lk}^{(n)}$ est égal à la probabilité de ne pas passer par E_k au cours de $n - r$ épreuves à partir de E_l et de ne pas passer par E_k au cours des r dernières épreuves. Comme après les $n - r$ premières épreuves l'état obtenu à partir de g_α appartiendra presque certainement à g_α la dernière probabilité sera $\leq 1 - w$ et, par suite, on aura

$$1 - w_{lk}^{(n)} \leq [1 - w_{lk}^{(n-r)}](1 - w)$$

d'où

$$1 - w_{lk}^{(n)} \leq [1 - w_{lk}^{(n-tr)}](1 - w)^t \leq (1 - w)^t.$$

Dès lors

$$q_{l0} = \lim_{n \to \infty} [1 - w_{lk}^{(n)}] = 0,$$

d'où $Q = 0$.

SECTION II.

CAS D'UNE SUITE CONTINUE D'ÉPREUVES.

Position du problème. — Restons encore dans l'hypothèse où le mouvement aléatoire qu'on étudie est celui d'un Système qui ne peut prendre qu'un nombre fini d'états possibles $E_1, E_2, \ldots, E_r$. Mais au lieu de le soumettre à une suite énumérable d'épreuves, supposons-le

assujetti à une suite d'épreuves qui se déroulent de façon continue dans le temps. Admettons donc qu'il y a une probabilité déterminée $P_{jk}(s, t)$ pour que le Système partant de l'état E_i à l'instant s arrive à l'état E_k à l'instant $t \geqq s$.

L'application des théorèmes des probabilités totales et composées montrera, comme à la page 14, que les $P_{jk}(s, t)$ vérifient le système de conditions.

$$(\mathrm{I}) \quad P_{jk}(s, t) = \sum_{i=1}^{i=r} P_{ji}(s, u) P_{ik}(u, t) \qquad \text{pour} \quad s \leqq u \leqq t, \ j, k = 1, 2, \ldots, r,$$

$$(\mathrm{T}) \qquad \sum_{k=1}^{k=r} P_{jk}(s, t) = 1 \qquad \text{pour} \quad s \leqq t, \ j = 1, \ldots, r$$

On aura, en outre, naturellement :

$$(\mathrm{P}) \qquad P_{jk}(s, t) \geqq 0 \qquad \text{pour} \quad s \leqq t, \ j, k = 1, \ldots, r$$

Mais il faut cette fois observer que $P_{jk}(s, s)$ ne peut être quelconque. Les états $E_1, \ldots, E_r$ étant supposés incompatibles, on aura nécessairement les conditions

$$(\mathrm{L}) \qquad P_{jk}(s, s) = \delta_{jk},$$

où

$$\delta_{jk} = \begin{cases} 1 & \text{si } j = k, \\ 0 & \text{si } j \neq k \end{cases}$$

Pour faciliter le problème du comportement asymptotique de $P_{jk}(s, t)$ quand $t \to +\infty$, nous commencerons d'abord par chercher quelle est la forme la plus générale des fonctions de s, t qui vérifient le système de conditions (I), (T), (P), (L).

A cet effet, nous chercherons d'abord les solutions les plus générales du système (I) et nous tiendrons compte successivement des conditions (L), (T), (P).

Mais il y aura lieu de préciser la nature analytique des fonctions de s et t que nous cherchons. La méthode de résolution et la forme générale des solutions correspondantes varieront en effet avec cette nature.

Existence de solutions discontinues. — Observons d'abord que nous connaissons déjà des solutions discontinues du problème.

Soient, en effet, les relations de récurrence

$$(1)\qquad a_{jk}^{(n)} = \sum_{i=1}^{r} a_{ji} a_{ik}^{(n-1)} = \sum_{i=1}^{r} a_{ji}^{(n-1)} a_{ik} \qquad (j, k = 1, \ldots, r),$$

avec

$$(2)\qquad a_{jk} \geqq 0, \qquad \sum_{k=1}^{r} a_{jk} = 1 \qquad (j, k = 1, \ldots, r)$$

$$(3)\qquad a_{jk}^{(0)} = \delta_{jk}$$

Si l'on se donne les a_{jk} satisfaisant aux conditions (2), les relations (1) et (3) déterminent de proche en proche les $a_{jk}^{(n)}$, et l'on aura

$$a_{jk}^{(n)} \geqq 0, \qquad \sum_{i=1}^{r} a_{jk}^{(n)} = 1$$

d'après le même raisonnement qu'à la page 38

On obtient alors le système de solutions discontinues annoncé plus haut, en posant

$$P_{jk}(s, t) = a_{jk}^{(Et-Es)} \qquad \text{pour} \quad s \leqq t$$

en appelant Ex la partie entière d'un nombre x (¹)

Il peut y avoir d'autres solutions discontinues. Pour simplifier le problème nous nous occuperons cependant surtout, dans ce qui va suivre, des solutions continues et plus particulièrement des solutions continues à variations bornées (déf., Premier Livre, note (¹), p. 273) et plus particulièrement encore des solutions dérivables à dérivées continues.

Nous passerons en revue, en les généralisant, les différentes méthodes utilisées par divers auteurs, méthodes qui ont chacune leur intérêt propre

Simple repérage du temps. — Faisons auparavant une observation. Supposons qu'on ait trouvé un système de fonctions $\varpi_{jk}(s, t)$ vérifiant les 4 conditions (I), (T), (P), (L), et soit $S = \Theta(s)$ une fonction con-

(¹) On observera que Et — Es est égal suivant les valeurs de t et de s, soit à E$(t-s)$, soit à $[1 + E(t-s)]$, de sorte qu'on ne peut pas dire rigoureusement que $P_{jl}(s, t)$ est dans ce cas une fonction de $t - s$: contrairement à ce qu'on pourrait penser, ce cas ne rentre pas exactement dans le cas homogène étudié page 248

tinue croissante de s; soit aussi $s = \Theta(S)$, sa fonction inverse. Il est clair que l'on aura encore une solution du problème, en prenant

$$P_{jk}(s, t) = \varpi_{jk}[\Theta(s), \Theta(t)]$$

Cela revient à dire que dans ce problème on n'a pas besoin de mesurer le temps t, il suffit de le repérer par un nombre $T = \Theta(t)$ qui est seulement assujetti à croître continuement avec le temps. Si cependant on veut pour les $P_{jk}(s, t)$ des fonctions soumises à certaines conditions de régularité, il faudra se restreindre à des familles correspondantes convenablement choisies de fonctions ϖ_{jk} et Θ. Par exemple, en prenant $\varpi_{jk}(s, t)$ et $\Theta(s)$ dérivables, il en sera de même de $\varpi_{jk}[\Theta(s), \Theta(t)]$.

PREMIÈRE MÉTHODE.

RÉDUCTION A UN SYSTÈME D'ÉQUATIONS DIFFÉRENTIELLES.

Méthode de Kolmogoroff. — Une première méthode consiste à ramener la résolution des équations fonctionnelles ci-dessus à l'intégration d'un système d'équations canoniques linéaires et homogènes. Le principe même de la méthode a pour effet de la réduire à la recherche des solutions dérivables.

Plus précisément, M. Kolmogoroff [1] suppose que $P_{jk}(s, t)$ est continu pour $s \leqq t$ et dérivable pour $s < t$, et il démontre que la dérivabilité à droite subsiste encore pour $s = t$. En effet, on a, pour $j = 1, 2, \ldots, r$, $s < t$, et $\Delta t > 0$

$$(4) \quad \frac{1}{\Delta t}[P_{jk}(s, t+\Delta t) - P_{jk}(s, t)] = \sum_{l} P_{jl}(s, t)\frac{P_{lk}(t, t+\Delta t) - P_{lk}(t, t)}{\Delta t},$$

en vertu des relations (I) et (L). Comme on le verra plus loin (p. 232), le déterminant $D(s, t)$ des $P_{jk}(s, t)$ est toujours $\neq 0$, on peut donc résoudre les équations (4) sous la forme

$$(5) \quad \frac{P_{lk}(t, t+\Delta t) - P_{lk}(t, t)}{\Delta t} = \sum_{j} \rho_{lj}(s, t)\frac{P_{jk}(s, t+\Delta t) - P_{jk}(s, t)}{\Delta t}.$$

Dès lors, quand $\Delta t \to 0$, puisque les seconds membres convergent, il en sera bien aussi de même des premiers.

On vient de démontrer qu'on définit des fonctions de t bien déterminées en posant

$$(6)\qquad U_{ik}(t) = \lim_{\Delta t \to 0} \frac{P_{ik}(t, t+\Delta t) - P_{ik}(t, t)}{\Delta t}$$

pour $\Delta t > 0$, c'est-à-dire que $P_{ik}(s, t)$ a pour $s = t$ au moins une dérivée à droite par rapport à t :

$$(7)\qquad U_{ik}(t) = \left[\frac{\partial^*}{\partial t} P_{ik}(s, t)\right]_{s=t}.$$

Il y a lieu d'observer, en outre, que ce résultat subsiste si l'on suppose seulement que les $P_{jk}(s, t)$ sont dérivables à droite pour $s < t$

Pour $i \neq k$, $P_{ik}(t, t) = 0$ et $\Delta t > 0$; la formule (6) montre donc que $U_{ik} \geqq 0$. D'autre part,

$$\sum_k U_{ik} = \lim_{\Delta t \to 0} \frac{\sum_k P_{ik}(t, t+\Delta t) - \sum_l P_{il}(t, t)}{\Delta t} = 0$$

Donc

$$(8)\qquad \sum_k U_{ik}(t) = 0,$$

et comme

$$U_{ik}(t) \geqq 0 \quad \text{pour } i \neq k,$$

on a

$$U_{ii}(t) \leqq 0$$

Enfin, d'après (4) et (7),

$$(9)\qquad \frac{\partial}{\partial t} P_{jk}(s, t) = \sum_i P_{ji}(s, t)\, U_{ik}(t) \qquad (j, k = 1, 2, \ldots, r)$$

Ces équations viennent d'être établies pour $t > s$. D'après (L) et (7), elles sont aussi satisfaites pour $t = s$, en ce qui concerne les dérivées à droite [1].

---

[1] La méthode actuelle ne s'applique pas à la dérivation à gauche. Mais la quatrième méthode développée plus loin permet de prouver (p. 245) que si les $P_{jk}(s, t)$ sont continues pour $s \leqq t$, on peut les prolonger pour $s > t$ et qu'alors si elles sont dérivables (à gauche comme à droite), pour $s < t$, elles sont aussi dérivables pour $s \geqq t$.

On verrait de même qu'on a pour $s < t$

$$\frac{\partial}{\partial s} P_{jk}(s, t) = -\sum_{i} U_{ji}(s) P_{ik}(s, t),$$

si les U_{ji} sont continues, ces équations sont aussi vérifiées pour $s = t$.

Si à un instant initial t_0, la probabilité initiale pour que le système se trouve à l'état E_j est $\varpi_j(t_0)$, on aura évidemment pour $t > t_0$

$$(10) \qquad \varpi_k(t) = \sum_{j} \varpi_j(t_0) P_{jk}(t_0, t)$$

Donc les $\varpi_k(t)$ sont dérivables, et en vertu de (9) et (10) vérifient le système

$$\frac{d}{dt} \varpi_k(t) = \sum_{i} \varpi_i(t) U_{ik}(t)$$

Celles-ci suffisent à déterminer les $\varpi_k(t)$ pour $t > t_0$, quand on se donne les valeurs initiales $\varpi(t_0)$.

Pour déduire des équations (9) le système de solutions dérivables le plus général du système de conditions (I), (T), (P), (L), M. Kolmogoroff a pu démontrer réciproquement que, si l'on se donne des fonctions $\alpha_{ik}(t)$ *continues* quelconques pourvu qu'elles vérifient les conditions

$$(11) \qquad \alpha_{ik}(t) \geqq 0 \qquad \text{pour} \quad i \neq k,$$

$$(12) \qquad \sum_{k} \alpha_{ik}(t) = 0,$$

alors, tout système de solutions $\varphi_{jk}(s, t)$ des équations

$$(13) \qquad \frac{\partial}{\partial t} \varphi_{jk}(s, t) = \sum_{i} \varphi_{ji}(s, t) \alpha_{ik}(t) \qquad (s \leqq t),$$

s'il vérifie les conditions initiales

$$(L') \qquad \varphi_{jk}(s, s) = \delta_{jk},$$

est un système de fonctions de s, t vérifiant les conditions (I), (T), (P), (L), c'est-à-dire que ces $\varphi_{ik}(s, t)$ peuvent être considérées comme les probabilités en chaîne d'un Système soumis à une suite continue d'épreuves.

On observera que toute solution $\varphi_{ik}(s, t)$, étant supposée dérivable en t pour que l'équation (13) ait un sens, est aussi continue en t comme les $\alpha_{ik}(t)$, que par suite les seconds membres de (13) sont continus. Donc la démonstration de la réciproque de M. Kolmogoroff concerne les solutions $\varphi_{ik}(s, t)$ des relations fonctionnelles (I), (T), (P), (L) qui sont dérivables avec des dérivées *continues*.

Démonstration de la réciproque. — Nous allons reproduire ici, en détaillant quelques points un peu délicats, la démonstration de M. Kolmogoroff.

Soit $\varphi_{ik}(s, t)$ un système de solutions de (13) vérifiant les conditions initiales (L′).

Démontrons d'abord le point le plus difficile, à savoir que les φ_{ik} vérifient la condition (P). Pour $s < t$, désignons par $\psi(s, t)$, ou plus brièvement par $\psi(t)$, la plus petite des valeurs de $\varphi_{ik}(s, t)$ quand i, k varient indépendamment par valeurs entières de 1 à r. Il y a au moins un couple de valeurs i, k (pouvant varier avec s et t) tel que l'on ait simultanément

$$\varphi_{ik}(s, t) = \psi(t) \qquad \text{et} \qquad \frac{\partial}{\partial t}\varphi_{ik}(s, t) = \frac{d^{+}\psi}{dt},$$

la deuxième dérivée par rapport à t étant prise à droite

En effet, pour s, t fixés, la première égalité aura lieu pour au moins un couple i, k, et en tout cas pour un système, g, de couples i, k. Si l'on se représente géométriquement les courbes $y = \varphi_{ik}(s, x)$, on verra intuitivement qu'il suffit pour réaliser les deux égalités de prendre dans le système g, un couple ik pour lequel le coefficient angulaire au point $x = t$ soit le plus petit possible (le plus grand possible s'il s'agissait de la dérivée à gauche). On peut présenter un raisonnement plus complet comme suit :

Soit λ une des limites finies ou infinies de $\psi(t + \Delta t)$ quand Δt tend vers zéro. On peut choisir une suite de valeurs de Δt : $\Delta_1 t, \ldots, \Delta_n t$, telle que

$$\lambda = \lim_{n \to \infty} \psi(t + \Delta_n t) \qquad \text{avec} \quad s < t + \Delta_n t$$

Mais il n'y a qu'un nombre fini de couples i, k et, pour chaque $\Delta_n t$, $\psi(t + \Delta_n t)$ est égal à l'un au moins des $\varphi_{ik}(s, t + \Delta_n t)$. Dès lors il existe au moins un de ces couples i_0, k_0, pour lequel l'égalité

$$\psi(t + \Delta_n t) = \varphi_{i_0 k_0}(s, t + \Delta_n t)$$

ait lieu une infinité de fois. Alors λ est l'une des limites de $\varphi_{i_0 k_0}(s, t+\Delta t)$ quand $\Delta t \to 0$, et comme $\varphi_{i_0 k_0}(s, t)$ est une fonction continue en t, on a $\lambda = \varphi_{i_0 k_0}(s, t)$. D'ailleurs, on a, d'après ce qui précède,

$$\varphi_{i_0 k_0}(s, t+\Delta t) = \psi(t+\Delta t) \leqq \varphi_{ik}(s, t+\Delta t)$$

pour une infinité de valeurs de Δt tendant vers zéro et quels que soient i, k. D'où

$$\varphi_{i_0 k_0}(s, t) \leqq \varphi_{ik}(s, t)$$

et, par suite,

$$\psi(t) = \varphi_{i_0, k_0}(s, t)$$

Ainsi toute limite λ de $\psi(t+\Delta t)$ est égale à $\psi(t)$ $\psi(t)$ est continue à droite, ce qui était intuitivement évident.

Soit maintenant μ l'une des limites, finie ou infinie, de $\dfrac{\psi(t+\Delta t)-\psi(t)}{\Delta t}$. C'est par exemple la limite quand Δt prend les valeurs $\Delta'_1 t$, $\ldots$, $\Delta'_n t$ qu'on peut supposer de même signe. Mais maintenant que nous savons $\psi(t)$ continue à droite, on pourrait recommencer le raisonnement précédent en prenant $\Delta_n t = \Delta'_n t$ On a donc un couple i_1, k_1 tel que, pour une infinité de valeurs de n,

$$\psi(t+\Delta'_n t) = \varphi_{i_1, k_1}(s, t+\Delta'_n t) \qquad \text{et} \qquad \psi(t) = \varphi_{i_1 k_1}(s, t),$$

d'où, pour ces valeurs de n, si, par exemple, $\Delta'_n t > 0$

$$\frac{\psi(t+\Delta'_n t)-\psi(t)}{\Delta'_n t} = \frac{\varphi_{i_1 k_1}(s, t+\Delta'_n t)-\varphi_{i_1, k_1}(s, t)}{\Delta'_n t}.$$

Par suite

$$\mu = \frac{\partial^*}{\partial t}\varphi_{i_1, k_1}(s, t) \qquad \text{avec} \quad \psi(t) = \varphi_{i_1 k_1}(s, t)$$

Ceci montre déjà que μ est fini et ne peut avoir que l'une des valeurs en nombre fini $\dfrac{\partial}{\partial t}\varphi_{ik}(s, t)$. Mais si $\dfrac{\psi(t+\Delta t)-\psi(t)}{\Delta t}$ avait une autre limite μ' que μ, on aurait

$$\mu' = \frac{\partial^*}{\partial t}\varphi_{i_2, k_2}(s, t), \qquad \psi(t) = \varphi_{i_2, k_2}(s, t),$$

d'où, pour une infinité de valeurs de n,

$$\frac{\psi(t+\Delta'_n t)-\psi(t)}{\Delta'_n t} \leqq \frac{\varphi_{i_2, k_2}(s, t+\Delta'_n t)-\varphi_{i_2, k_2}(s, t)}{\Delta'_n t}$$

et, par suite,

$$\mu \leqq \frac{\partial^*}{\partial t}\varphi_{i_2, k_2}(s, t) = \mu'.$$

On verrait de même que $\mu' \leqq \mu$; d'où $\mu = \mu'$; ainsi $\psi(t)$ est dérivable à

droite et il existe un couple i_1, k_1 tel qu'on ait simultanément

$$\psi(t) = \varphi_{i_1 k_1}(s, t), \qquad \frac{d^*\psi(t)}{dt} = \frac{\partial^*}{\partial t}\varphi_{i_1 k_1}(s, t)$$

Dès lors si $\psi(t)$ était négatif, on aurait, en vertu des hypothèses sur les α,

$$\alpha_{jk}(t)\varphi_{ij}(s, t) \geqq \alpha_{jk}(t)\psi(t) \qquad \text{pour} \quad j \neq k$$

et

$$\alpha_{kk}(t)\varphi_{ik}(s, t) \geqq 0;$$

d'où

$$\frac{d^*\psi(t)}{dt} = \frac{\partial^*}{\partial t}\varphi_{ik}(s, t) = \sum_j \alpha_{jk}(t)\varphi_{ij}(s, t) \geqq \left[\sum_{j \neq k} \alpha_{jk}(t)\right]\psi(t) = R(t)\psi(t)$$

d'où, enfin, pour la dérivée à droite,

$$\frac{d^*\psi(t)}{dt} - R(t)\psi(t) \geqq 0 \tag{14}$$

$R(t)$ désignant une fonction continue $\geqq 0$. En prenant les $\Delta'_n t < 0$ [*voir* note (¹), p. 205], on prouverait de même que $\psi(t)$ est dérivable à gauche Et comme on a vu plus haut que la dérivée à gauche est au moins égale à la dérivée à droite, les deux dérivées sont $\geqq R(t)\psi(t)$.

Or, en vertu de (L′), $\psi(s) = 0$. De cette égalité et de (14) on conclut que $\psi(t)$ ne peut être négatif pour $t > s$, car on a pour la dérivée à droite comme pour la dérivée à gauche

$$\frac{d}{dt}\left[\psi(t)e^{-\int_s^t R(\tau)\,d\tau}\right] = e^{-\int_s^t R(\tau)\,d\tau}\left\{\frac{d\psi}{dt} - \psi(t)R(t)\right\} \geqq 0$$

Donc le crochet est une fonction non décroissante de t qui est nulle pour $t = s$, dès lors $\psi(t) \geqq 0$ et, par suite, tous les $P_{ik}(s, t)$ sont $\geqq 0$.

Les conditions initiales imposées aux $\varphi_{jk}(s, t)$ expriment que la condition (L′) est réalisée. D'autre part

$$\frac{\partial}{\partial t}\left[\sum_k \varphi_{jk}(s, t)\right] = \sum_i \varphi_{ji}(s, t)\left[\sum_k \alpha_{ik}(s, t)\right] = 0$$

d'après (12), donc $\sum_k \varphi_{jk}(s, t)$ est constant pour $t \geqq s$, et, par suite, égal à 1, en vertu de (L′).

Enfin, en posant, pour $s \leqq u$, $\theta_{jk}(s, t) = \varphi_{jk}(s, t)$ si $s \leqq t \leqq u$, et

$\theta_{jk}(s, t) = \sum_{i} \varphi_{ji}(s, u)\varphi_{ik}(u, t)$ si $s \leqq u \leqq t$, les fonctions $\theta_{jk}(s, t)$ sont continues et l'on s'assure, par substitution, qu'elles vérifient les équations différentielles (13). Ce sont des solutions continues qui ont les mêmes valeurs initiales, pour $t = u$, que les solutions $\varphi_{jk}(s, t)$. Elles leur sont donc identiques, ce qui revient à dire que les $\varphi_{jk}(s, t)$ vérifient l'équation fonctionnelle (I'). En resumé, les $\varphi_{jk}(s, t)$ vérifient les quatre conditions (I'), (T'), (L'), (P') ([1]).

Extension de la méthode précédente. — Nous allons modifier la méthode de M. Kolmogoroff (sans en changer la marche générale) de façon à étendre le champ des solutions envisagées. De plus, nous allons, pour commencer, ne pas tenir compte des conditions (P), (T), (L) et nous limiter, d'abord, à l'étude de l'équation d'itération (I). Pour mettre en évidence qu'il ne s'agit plus nécessairement de probabilités, nous remplacerons encore les $P_{jk}(s, t)$ par $\varphi_{jk}(s, t)$, et nous considérerons l'équation d'itération

$$(\mathrm{I}') \qquad \varphi_{jk}(s, t) = \sum_{i=1}^{r} \varphi_{ji}(s, u)\, \varphi_{ik}(u, t) \qquad (s \leqq u \leqq t), \ (j, k = 1, \ldots, r)$$

Nous allons en chercher les solutions qui sont des fonctions de t *continues et à variations bornées;* nous supposerons de plus que le déterminant $D(s, t)$ des $\varphi_{jk}(s, t)$ ne s'annule jamais ([2]) pour $t \geqq s$.

Il en résulte encore que l'on peut résoudre les équations (I') obtenues pour k fixe, $j = 1, \ldots, r$ sous la forme

$$(15) \qquad \varphi_{ik}(u, t) = \sum_{j} \psi_{ij}(s, u)\, \varphi_{jk}(s, t),$$

où les $\psi_{ij}(s, u)$ sont des fonctions continues de u.

([1]) Il ne saurait y avoir ici de confusion entre la condition $\sum_{i} p_{ik} = 1$ désignée par (T') dans la Section I et la condition $\sum_{k} \varphi_{ik}(s, t) = 1$ que nous avons appelée (T') dans la Section II, par analogie avec les conditions (L'), (P').

([2]) Cette dernière supposition est vérifiée d'elle-même dans l'application aux probabilités, en ce qui concerne les solutions continues, comme on le verra, page 214

Dès lors, si $s = t_0 \leqq t_1 \leqq t_2 \ldots \leqq t_n = t$, on pourra écrire, en remplaçant ci-dessus s par un nombre quelconque $s' \leqq s$ et u par s,

$$(16)\quad \varphi_{ik}(s,t) - \varphi_{ik}(s,s) = \sum_{\alpha=1}^{n} [\varphi_{ik}(s,t_\alpha) - \varphi_{ik}(s,t_{\alpha-1})]$$

$$= \sum_{l} \sum_{\alpha=1}^{n} \psi_{il}(s',s)[\varphi_{lk}(s',t_\alpha) - \varphi_{lk}(s',t_{\alpha-1})]$$

D'autre part, si $s = t_0 \leqq t_1 \ldots \leqq t_n = t$, on a, d'après (I'),

$$\varphi_{jk}(s,t_{\alpha+1}) - \varphi_{jk}(s,t_\alpha) = \sum_{i} \varphi_{ji}(s,t_\alpha)[\varphi_{ik}(t_\alpha,t_{\alpha+1}) - \varphi_{ik}(t_\alpha,t_\alpha)]$$

d'où

$$(17)\quad \varphi_{jk}(s,t) - \varphi_{jk}(s,s) = \sum_{i} \sum_{\alpha=0}^{n-1} \varphi_{ji}(s,t_\alpha)[\varphi_{ik}(t_\alpha,t_{\alpha+1}) - \varphi_{ik}(t_\alpha,t_\alpha)]$$

et, en vertu de (15), le second membre s'écrit

$$= \sum_{i} \sum_{\alpha=0}^{n-1} \varphi_{ji}(s,t_\alpha) \sum_{l} \psi_{il}(s',t_\alpha)[\varphi_{lk}(s',t_{\alpha+1}) - \varphi_{lk}(s',t_\alpha)]$$

où $s' \leqq s$.

Comme $\varphi_{ji}(s,t)$, $\psi_{il}(s,t)$ sont des fonctions continues de t et $\varphi_{lk}(s,t)$ est à variation bornée en t, on a à la limite au second membre une somme d'intégrales de Stieltjes (intégrales définies par exemple dans ce Traité, vol. I, fasc. 1, p. 111). D'où

$$(18)\quad \varphi_{jk}(s,t) - \varphi_{jk}(s,s) = \sum_{i} \sum_{l} \int_s^t \varphi_{ji}(s,\tau)\,\psi_{il}(s',\tau)\,d\varphi_{lk}(s',\tau).$$

D'ailleurs, en prenant s' fixe, s variable, $s' \leqq s \leqq t$, la fonction

$$(19)\quad V_{ik}(t) = \sum_{l} \int_s^t \psi_{il}(s',\tau)\,d\varphi_{lk}(s',\tau)$$

est une somme d'intégrales de Stieltjes qui est une fonction de t continue à variation bornée, et l'on peut écrire

$$(20)\quad \varphi_{jk}(s,t) - \varphi_{jk}(s,s) = \sum_{i} \int_s^t \varphi_{ji}(s,\tau)\,dV_{ik}(\tau).$$

Les fonctions $\varphi_{1i}(s, t), \ldots, \varphi_{ji}(s, t)$ sont donc, pour s fixe, des solutions d'un système de la forme

$$(21) \qquad y_k(t) = y_k(s) + \sum_{i=1}^{j} \int_s^t y_i(\tau)\, dV_{ik}(\tau) \qquad (k = 1, \ldots, j),$$

où les $V_{ik}(t)$ sont des fonctions continues à variations bornées [1].

C'est le système qui généralise le système d'équations différentielles de M. Kolmogoroff. A ce dernier se ramène notre système (21) quand on se restreint aux solutions dérivables en t, cas où les $V_{ik}(t)$ sont des fonctions admettant des dérivées continues $U_{ik}(t)$. On a alors

$$y_k(t) = y_k(s) + \sum_i \int_s^t y_i(\tau)\, U_{ik}(\tau)\, d\tau$$

et en dérivant, ce qu'on a le droit de faire,

$$\frac{dy_k}{dt} = \sum_i U_{ik}(t)\, y_i.$$

C'est le système de M. Kolmogoroff, qui, d'ailleurs, se trouve établi sans supposer qu'il s'agit de probabilités.

Dans le cas des probabilités, que peut-on dire sur les $V_{ik}(t)$?

En vertu de (15), on a

$$\sum_\alpha [\varphi_{ik}(t_\alpha, t_{\alpha+1}) - \varphi_{ik}(t_\alpha, t_\alpha)] = \sum_l \sum_\alpha \varphi_{il}(s', t_\alpha)\, [\varphi_{lk}(s', t_{\alpha+1}) - \varphi_{lk}(s', t_\alpha)],$$

d'où, par (19),

$$V_{ik}(t) = \lim \left[\sum_\alpha \varphi_{ik}(t_\alpha, t_{\alpha+1}) - \varphi_{ik}(t_\alpha, t_\alpha)) \right].$$

D'après la condition (T), on en déduit que

$$(22) \qquad \sum_k V_{ik}(t) = 0$$

---

(1) A vrai dire, ceci ne serait établi que pour s et t au moins égaux à un nombre fixe s' dont dépendraient les $V_{ik}(t)$. Mais la quatrième Méthode montre (p. 245) que cette restriction peut être levée.

et, d'après la condition (L), on a, pour $i \neq k$,

$$V_{ik}(t) = \lim \sum_{\nu} \varphi_{ik}(t_\nu, t_{\nu+1})$$

$V_{ik}(t)$ est donc pour $i \neq k$ une fonction $\geqq 0$ et même non décroissante quand t croît. Il résulte alors de (22) que $V_{kk}(t)$ est, au contraire, une fonction $\leqq 0$ et non croissante de t.

Extension intermédiaire. — Considérons le problème qui consiste à trouver les solutions « absolument continues » (défin., Premier Livre, p. 33) sur un segment S, de l'équation fonctionnelle I', c'est-à-dire les solutions représentables par une intégrale. Elles sont, en particulier, continues et à variation bornée.

On peut donc leur appliquer les raisonnements qui viennent d'être développés. Seulement, ici, on peut représenter φ_{ik} sous la forme

$$\varphi_{ik}(s, \tau) = \int_s^\tau \theta_{ik}(s, \tau_1)\, d\tau_1 + \varphi_{ik}(s, s),$$

d'où, par (19),

$$(23) \qquad V_{ik}(t) = \int_s^t \sum_l \varphi_{il}(s', \tau)\, \theta_{lk}(s', \tau)\, d\tau = \int_s^t U_{ik}(\tau)\, d\tau$$

et

$$(24) \qquad \varphi_{ik}(s, t) - \varphi_{ik}(s, s) = \sum_l \int_s^t \varphi_{il}(s, \tau)\, U_{lk}(\tau)\, d\tau,$$

où les $U_{ik}(t)$ sont des fonctions sommables.

C'est le système intégral qui remplace à la fois dans le cas actuel le système différentiel de M. Kolmogoroff et le système intégral (20). Ce dernier comprenait des intégrales de Stieltjes, le système (24) comprend des intégrales de Lebesgue.

Dans le cas des probabilités, on a,

$$0 = \sum_k V_{ik}(t) = \int_s^t \sum_k U_{ik}(\tau)\, d\tau,$$

pour tout t sur S ; on a donc sur S

$$(25) \qquad \sum_k U_{ik}(t) = 0.$$

D'autre part, puisque, pour $i \neq k$, $V_{ik}(t)$ est non décroissante, $U_{ik}(t)$ étant presque partout la dérivée de $V_{ik}(t)$ et pouvant être ailleurs pris égal à zéro, sera partout $\leqq 0$ sur S.

DEUXIÈME ET TROISIÈME MÉTHODES.

SOLUTION SOUS FORME D'UNE SÉRIE D'INTÉGRALES MULTIPLES D'ORDRES CROISSANTS.

Les deux méthodes de M. Hostinský. — La méthode de M. Kolmogoroff ne fournit pas l'expression des solutions. Pour combler cette lacune, M. Hostinský [14bis, 17] a proposé successivement deux méthodes en 1932 et en 1937.

La première est fondée sur l'ingénieux emploi de la notion d'intégrale d'une substitution (due à M. Volterra [1]) et appliquée directement à l'équation fonctionnelle (I). Elle en fournit une solution très générale sans qu'on sache si elle donne toutes les solutions.

La seconde méthode que nous allons reproduire a l'avantage didactique d'utiliser un procédé plus connu et l'avantage scientifique de donner *toutes* les solutions à dérivées continues. Elle consiste a partir des équations différentielles de M. Kolmogoroff et à les intégrer par la méthode des approximations successives. (Nous rappelons celle-ci dans la Note E, p. 291).

Les deux méthodes aboutissent cependant à la même formule (29).

Au lieu de nous imposer tout de suite les conditions (P), (T), contentons-nous d'abord de chercher, avec M. Hostinský, les solutions de l'équation d'itération (I') qui vérifient les conditions initiales

$$\text{(L')} \qquad \varphi_{jk}(s, s) = \delta_{jk}$$

Et parmi celles-ci, adressons-nous d'abord (c'est le cas où s'appliquent les résultats de M. Hostinský) aux solutions de (I') qui sont dérivables avec des dérivées continues par rapport à l'ensemble des deux variables s, t (en supposant toujours $s \leqq t$). On a

(pour $s \leqq t \leqq v$)

$$\varphi_{jk}(s, v) = \sum_i \varphi_{ji}(s, t)\, \varphi_{ik}(t, v), \tag{I'}$$

d'où

$$\frac{\partial}{\partial v} \varphi_{jk}(s, v) = \sum_i \varphi_{ji}(s, t) \left[\frac{\partial}{\partial v} \varphi_{ik}(t, v)\right]$$

Les hypothèses faites montrent qu'en faisant tendre v vers t, on obtient

$$\frac{\partial}{\partial t} \varphi_{jk}(s, t) = \sum_i \varphi_{ji}(s, t)\, U_{ik}(t), \tag{26}$$

où

$$U_{ik}(t) = \left[\frac{\partial}{\partial v} \varphi_{ik}(t, v)\right]_{v=t}$$

est une fonction continue de t.

Dès lors, pour chaque valeur de j, $\varphi_{j1}(s, t), \ldots, \varphi_{jn}(s, t)$ sont des fonctions de t égales à $\delta_{j1}, \ldots, \delta_{jn}$ pour $t = s$ et vérifiant pour $t \geqq s$ le système d'équations différentielles

$$\frac{dx_k}{dt} = \sum_i^n x_i U_{ik}(t), \qquad (k = 1, \ldots, n) \tag{27}$$

Or, la méthode des approximations successives nous apprend que, les fonctions $U_{ik}(t)$ étant continues, ce système n'a qu'un seul système de solutions prenant des valeurs donnees pour $t = s$. Et pour chaque valeur de j, M. Hostinský [17] a fait observer (Note E, p. 291) qu'une application immédiate de cette même méthode conduit à un développement simple des solutions de ce système égales aux δ_{jk} pour $t = s$, donc à un développement simple des $\varphi_{jk}(s, t)$, soit

$$\begin{aligned} \varphi_{jk}(s, t) = \delta_{jk} &+ \int_s^t U_{jk}(\tau_1)\, d\tau_1 + \sum_i \int_s^t \int_s^{\tau_2} U_{ji}(\tau_1)\, U_{ik}(\tau_2)\, d\tau_1\, d\tau_2 + \ldots \\ &+ \sum_{i_1 \ldots i_{n-1}} \int_s^t \int_s^{\tau_n} \cdots \int_s^{\tau_2} U_{ji_1}(\tau_1) U_{i_1 i_2}(\tau_2) \ldots U_{i_{n-1}k}(\tau_n)\, d\tau_1 \ldots d\tau_n + \ldots, \end{aligned} \tag{28}$$

développement dont la même méthode permet de prouver qu'il est normalement convergent (¹).

(¹) Une série dont les termes dépendent de paramètres variant dans un domaine D est, selon Baire, normalement convergente sur D si ses termes sont majorés en valeurs absolues par une série convergente à termes constants.

Ainsi toute solution vérifiant (L′) de l'équation d'itération (I′), qui est dérivable en t, avec une dérivée continue par rapport à l'ensemble de s et de t, est nécessairement développable sous cette forme où les $U_{jk}(t)$ sont des fonctions continues.

Réciproquement, donnons-nous arbitrairement des fonctions continues $\alpha_{ik}(t)$ et formons les expressions

$$(29)\quad \Phi_{jk}(s, t) = \delta_{jk} + \int_s^t \alpha_{jk}(\tau_1)\, d\tau_1 + \ldots$$
$$+ \sum_{i_1 \ldots i_{n-1}} \int_s^t \int_s^{\tau_n} \cdots \int_s^{\tau_2} \alpha_{j i_1}(\tau_1) \ldots \alpha_{i_{n-1} k}(\tau_n)\, d\tau_1 \ldots d\tau_n + \ldots$$

En majorant cette série comme dans la méthode des approximations successives (Note E), on voit qu'elle est normalement convergente. Les $\Phi_{jk}(s, t)$ sont alors évidemment des fonctions continues égales aux δ_{jk} pour $t = s$. On voit de même que la série des dérivées en t de $\Phi_{jk}(s, t)$ a ses termes continus en s et t et est normalement convergente. Par suite, les séries (29) représentent des fonctions $\Phi_{jk}(s, t)$ dérivables en t et dont les dérivées en t sont continues par rapport à l'ensemble de s et de t.

Enfin, ce sont des solutions de (I′). On pourrait sans doute le voir par une manipulation d'intégrales multiples après substitution dans (I′). On peut aussi observer que ces $\Phi_{jk}(s, t)$ sont, pour j et s fixes, solutions des équations

$$\frac{dx_k}{dt} = \sum_i x_i(t)\, \alpha_{ik}(t).$$

Les solutions sont, comme on l'a vu, de la forme

$$x_k(t) = \sum_i x_i(u)\, \Phi_{ik}(u, t).$$

Appliquons en particulier aux solutions

$$x_k(t) = \Phi_{jk}(s, t),$$

on aura

$$\Phi_{jk}(s, t) = \sum_i \Phi_{ji}(s, u)\, \Phi_{ik}(u, t).$$

En résumé, la solution $\Phi_{jk}(s, t)$, la plus générale, dérivable

en t, avec une dérivée en t continue par rapport à l'ensemble de s et de t du système (I'), (L'), *est de la forme* (28), *où les* $\alpha_{ik}(t)$ *sont des fonctions continues arbitraires de t.*

C'est là le résultat, indépendant de la théorie des probabilités et de l'intégration des substitutions, qui a été obtenu par M. Hostinský dans sa seconde méthode.

Première extension. — Ce résultat peut même être transformé de façon à s'appliquer à un cas plus général, en faisant usage d'une part de l'extension donnée plus haut à la méthode de M. Kolmogoroff et, d'autre part, de l'application (*voir* Fréchet [199]) de la méthode des approximations successives à une forme d'équations un peu plus générale que les équations différentielles linéaires.

En effet, on a vu, page 213, que tout système de solutions absolument continues, $\varphi_{jk}(s, t)$, des équations (I'), (L') vérifie, pour j fixe, des équations de la forme (27), mais dans lesquelles les coefficients $U_{ik}(t)$ ne sont plus necessairement continus et sont simplement sommables sur S. La méthode des approximations successives montre alors (Note E) qu'il existe, pour chaque j, un système de solutions et un seul de ces équations, prenant, pour $t = s$, les valeurs $\varphi_{jk}(s, s) = \delta_{jk}$. Ce système de solutions est encore fourni par la série (28) normalement convergente sur S. Ces solutions sont absolument continues sur S.

Réciproquement, si l'on se donne arbitrairement des fonctions $\alpha_{ik}(t)$ sommables sur S, les séries (29) seront normalement convergentes sur S vers des fonctions absolument continues et seront des solutions de (I')

En résumé, *la solution absolument continue la plus générale, du système* (I'), (L') *est de la forme* (29), *où les* $\alpha_{ik}(t)$ *sont des fonctions sommables arbitraires de t.*

Seconde extension. — On a vu page 210 que tout système de solutions continues et à variation bornée sur S des équations (I'), (L'), vérifie pour j fixe des équations de la forme

$$y_k(t) = y_k(s) + \sum_{i=1}^{r} \int_s^t y_i(\tau)\, dV_{ik}(\tau), \tag{30}$$

où les $V_{ik}(\tau)$ sont nulles pour $\tau = s$, continues et à variations bornées. D'autre part (Note E), un tel système n'admet qu'un système de solutions prenant pour $t = s$, les valeurs $\delta_{j1}, \ldots, \delta_{jn}$, et ces solutions peuvent être développées en série de façon simple. De sorte que l'on obtient ainsi le développement simple suivant des φ_{jk}

$$(31)\quad \varphi_{jk}(s,t) = \delta_{jk} + V_{jk}(t) + \int_s^t \sum_i V_{ji}(\tau)\,dV_{ik}(\tau) + \int_s^t\int_s^{\tau_n}\cdots\int_s^{\tau_2} \sum_{i_1,\ldots,i_n} V_{ji_1}(\tau_1)\,dV_{i_1i_2}(\tau_2)\ldots dV_{i_nk}(\tau_{n+1}) + \ldots$$

où les intégrations successives doivent être entendues au sens de Stieltjes, la série étant normalement convergente sur S.

Réciproquement, soient $V_{ik}(t)$ des fonctions continues et à variations bornées sur S, nulles pour $t = s$, choisies arbitrairement et formons les séries

$$(32)\quad \Phi_{jk}(s,t) = \delta_{jk} + V_{jk}(t) + \ldots + \int_s^t\int_s^{\tau_n}\cdots\int_s^{\tau_2} \sum_{i_1,\ldots,i_n} V_{ji_1}(\tau_1)\,dV_{i_1i_2}(\tau_2)\ldots dV_{i_nk}(\tau_{n+1}) + \ldots$$

On voit facilement que ces séries convergent normalement et représentent des fonctions de t égales aux δ_{jk} pour $t = s$ et continues et à variations bornées.

Elles vérifient le système

$$\Phi_{jk}(s,t) = \delta_{jk} + \int_s^t \sum_i \Phi_{ji}(s,\tau)\,dV_{ik}(\tau).$$

Partant de ces équations, le raisonnement employé, page 216, subsiste dans le cas actuel plus général et montre que ces fontions Φ_{jk} vérifient l'équation fonctionnelle

$$\Phi_{jk}(s,t) = \sum_i \Phi_{ji}(s,u)\,\Phi_{ik}(u,t).$$

En résumé, *la solution* $\Phi_{jk}(s,t)$ *de* (I′), (L′) *la plus générale parmi celles qui sont des fonctions de t continues et à variations bornées sur le segment* S *est de la forme* (32), *où les* $V_{ik}(t)$ *sont des fonctions continues et à variations bornées sur* S, *nulles pour* $t = s$ *et choisies arbitrairement.*

Ce sont là deux nouvelles extensions — indépendantes de la théorie des probabilités et de l'intégration des substitutions — et qui fournissent les solutions appartenant à une famille de fonctions plus générales que celles de M. Hostinský.

QUATRIÈME MÉTHODE

SOLUTION EN TERMES FINIS.

Caractère de la solution. — Après la publication de la méthode de M. Kolmogoroff en 1931 et de la première méthode de M. Hostinský en 1932, le problème de la détermination explicite de la solution continue la plus générale des équations fonctionnelles (I'), (L'), pages 206 et 210, restait ouvert. Une méthode simple et directe nous a permis (Fréchet, [150]) de résoudre ce problème en 1932. La seconde méthode de M. Hostinský (en 1937) a permis ensuite d'obtenir aussi, d'une façon et sous une forme différente, les solutions, à dérivées continues, les plus générales, comme nous l'avons vu plus haut.

Les modifications que nous avons indiquées plus haut, à cette seconde méthode, permettent même d'obtenir des classes plus générales de solutions.

Malgré cela, notre méthode garde encore l'avantage essentiel de fournir *une expression beaucoup plus simple de la solution.* Elle conduit, comme la méthode de M. Hostinský, à exprimer la solution en fonction du même nombre r^2 de fonctions arbitraires (arbitraires à des conditions près de régularité). Mais, tandis que la méthode de M. Hostinský exprimait la solution sous forme d'une série *illimitée* d'intégrales multiples d'ordres *croissants*, où les fonctions arbitraires étaient combinées entre elles comme des polynomes de degré *croissant* indéfiniment, notre solution fait intervenir des fonctions arbitraires sous forme finie, et même, plus précisément, s'exprime en fonctions rationnelles de ces fonctions arbitraires. Malgré tout, nous avons tenu à exposer les quatre méthodes qui sont d'esprits différents et qui ont chacune leur intérêt propre.

Étude de l'equation fonctionnelle (F_r). — Pour mettre en évidence que nous faisons d'abord abstraction des conditions (P'), (T'), changeons encore de notations et proposons-nous la détermination de *toutes les solutions continues*, soumises à la condition

$$(L') \qquad \varphi_{jk}(s, s) = \begin{cases} 1 & \text{si } j = k, \\ 0 & \text{si } j \neq k \end{cases}$$

de l'équation fonctionnelle à r termes (F'), où ici

$$(F_r) \qquad \varphi_{jk}(s, t) = \sum_{l=1}^{l=r} \varphi_{jl}(s, u)\varphi_{lk}(u, t) \qquad (s \leqq u \leqq t).$$

La relation (F_r) a au moins le système de solutions obtenu en prenant tous les $\varphi_{ik}(s, t)$ nuls. Nous laisserons toujours de côté cette solution. Il y a encore en évidence d'autres solutions, par exemple le système de solutions où tous les $\varphi_{jk}(s, t)$ sont égaux à $\frac{1}{r}$. Considérons le système le plus général de solutions non toutes nulles, nous allons voir que deux cas se présentent. Pour cela, désignons par $D(s, t)$ le déterminant d'un système de solutions $\varphi_{jk}(s, t)$. On aura, en vertu de (F_r),

$$(F_1) \qquad D(s, t) = D(s, u)D(u, t) \qquad \text{pour } s \leqq u \leqq t.$$

Or, observons que cette relation auxiliaire (F_1) est celle à laquelle se réduit l'équation fonctionnelle (F_r) dans le cas où $r = 1$, cas où $D(s, t)$ se réduit à $\varphi_{11}(s, t)$. Il sera donc utile de résoudre d'abord cette équation fonctionnelle, d'abord parce que la discussion qui en résulte nous éclairera sur les valeurs du déterminant des φ_{jk} pour $r \neq 1$, ensuite parce que la résolution du cas $r = 1$ nous guidera vers la solution du cas $r \geqq 1$.

Remarque sur la condition (L'). — Faisons auparavant une remarque. On a

$$\varphi_{jk}(u, u) = \sum_l \varphi_{jl}(u, u)\varphi_{lk}(u, u).$$

Considérons les équations obtenues pour k fixe, $j = 1, \ldots, r$:

$$\varphi_{j1}(u, u)\varphi_{1k}(u, u) + \ldots$$
$$+ \varphi_{jk}(u, u)[\varphi_{kk}(u, u) - 1] + \ldots + \varphi_{jr}(u, u)\varphi_{rk}(u, u) = 0.$$

On peut les envisager comme des équations linéaires et homogènes en $\varphi_{1k}(u, u), \ldots, \varphi_{kk}(u, u) - 1, \ldots, \varphi_{rk}(u, u)$, dont le déterminant des coefficients est $D(u, u)$. Donc, si $D(u, u) \neq 0$, on a

$$(L_1) \qquad \varphi_{jk}(u, u) = \begin{cases} 1 & \text{si } j = k, \\ 0 & \text{si } j \neq k, \end{cases}$$

conditions analogues aux conditions (L'). Réciproquement si les conditions (L') ont lieu, $D(u, u) \neq 0$. Ainsi *les conditions* (L') *sont entièrement équivalentes à la condition que* $D(u, u) \neq 0$, c'est-à-dire que le déterminant $D(s, t)$ des solutions φ_{jl} de (F_r) est $\neq 0$ pour $s = t = u$.

Solution continue la plus générale de l'equation fonctionnelle (F_1) — Soit à résoudre l'équation fonctionnelle

$$(F_1) \qquad D(s, t) = D(s, u)\, D(u, t) \qquad \text{pour} \quad s \leqq u \leqq t$$

Nous déterminerons plus loin les solutions continues du problème. Mais pour commencer, *cherchons d'abord les solutions* du problème *qui ne sont jamais nulles*. On les obtient immédiatement. Soit en effet u_0 un point quelconque. On a les cas suivants :

$$D(s, t) = \frac{D(s, u_0)}{D(t, u_0)} \qquad \text{pour} \quad s \leqq t \leqq u_0,$$

$$D(s, t) = D(s, u_0)\, D(u_0, t) \qquad \text{pour} \quad s \leqq u_0 \leqq t,$$

$$D(s, t) = \frac{D(u_0, t)}{D(u_0, s)} \qquad \text{pour} \quad u_0 \leqq s \leqq t$$

on peut donc écrire, dans tous ces cas,

$$(1) \qquad D(s, t) = \frac{A(t)}{A(s)} \qquad \text{pour} \quad s \leqq t,$$

en posant

$$(2) \qquad A(s) = D(u_0, s) \qquad \text{pour} \quad u_0 \leqq s,$$

$$(3) \qquad A(s) = \frac{1}{D(s, u_0)} \qquad \text{pour} \quad s \leqq u_0.$$

Dans ces formules, $A(s)$ et $A(t)$ sont toujours, par hypothèse, finis et $\neq 0$.

C'est d'ailleurs la solution toujours finie et jamais nulle de (F_1) *la plus générale*, si l'on prend inversement pour $A(s)$ une fonction

entièrement arbitraire — pourvu que cette fonction $A(s)$ soit partout *finie et* $\neq 0$ — en vertu de l'identité

$$\frac{A(t)}{A(s)} = \frac{A(t)}{A(u)}\,\frac{A(u)}{A(s)}.$$

On notera que la fonction (1) est définie et solution de (F_1), non seulement dans l'hypothèse $s < t$ que nous nous étions imposé mais aussi pour $s > t$.

Ce qui précède subsiste d'ailleurs si l'on suppose $D(s, t)$ défini (et toujours $\neq 0$) seulement à l'*intérieur* d'un segment (S, T) ou d'une demi-droite $(-\infty, T)$ ou $(S, +\infty)$.

Que $D(s, t)$ soit continu ou non, puisse ou non s'annuler, on observe encore que

$$D(t, t) = D(t, t)\,D(t, t).$$

Donc $D(t, t)$ ne peut jamais prendre que les valeurs 0 ou 1. D'autre part, on a évidemment

$$D(s, t) = D(s, u)\,D(u, u)\,D(u, t) = D(u, u)\,D(s, t),$$

si $s \leqq u \leqq t$. Dès lors, pour tout point u d'un intervalle (s, t) tel que $D(s, t) \neq 0$, on a $D(u, u) = 1$.

De tout ce qui précède, on va déduire les solutions continues de (F_1). Nous laisserons de côté la solution continue évidente $D(s, t) \equiv 0$. Pour toute solution $D(s, t)$ continue, la fonction continue $D(t, t)$ ne pouvant prendre que les valeurs 0 ou 1 gardera constamment la même valeur. Si la solution continue $D(s, t)$ n'est pas identiquement nulle, il y aura au moins un couple s, t tel que $D(s, t) \neq 0$ et au moins un point u appartenant à (s, t) tel que $D(u, u) = 1$. Il en résulte que $D(t, t) = 1$ pour toute valeur de t.

Mais alors $D(s, t)$ reste toujours positif. Car s'il existait un couple s, t tel que $D(s, t) \leqq 0$, on aurait nécessairement $s \neq t$, et non $s = t$, et si u est le milieu de (s, t) on aurait

$$D(s, u)\,D(u, t) \leqq 0;$$

donc

$$D(s, u) \leqq 0 \quad \text{ou} \quad D(u, t) \leqq 0.$$

On formerait alors une suite d'intervalles — chacun (s_n, t_n) contenu dans le précédent (s_{n-1}, t_{n-1}) et de longueur moitié — pour lesquels

$D(s_n, t_n) \leqq 0$. Il y aurait un point limite t pour lequel on aurait $D(t', t') \leqq 0$ et non $D(t', t') = 1$. D'où la contradiction annoncée.

Enfin, quand $D(s, t)$ est continue, non $\equiv 0$ et donnée, l'expression de $A(s)$ donnée par les formules (2) et (3) sera positive et continue. Réciproquement, si $A(s)$ est une fonction toujours $\neq 0$ et donnée arbitrairement, alors pour que le quotient $D(s, t) = \frac{A(t)}{A(s)}$ soit continu par rapport à l'ensemble (s, t), il suffit évidemment et de plus il faut que $A(t)$ soit continue. Car si t_0 est pris arbitrairement, on peut prendre $s_0 < t_0$, et alors, pour $t > s_0$,

$$A(t) = D(s_0, t)\, A(s_0)$$

est continue près de t_0 qui est arbitraire. De plus $A(t)$ étant continue et $\neq 0$ garde un signe constant; alors, en écrivant, au besoin,

$$D(s, t) = \frac{-A(t)}{-A(s)},$$

on pourra supposer que $A(s)$ reste > 0.

En résumé, toute solution continue $D(s, t)$ de l'équation (F_1) ou bien est $\equiv 0$, ou bien est constamment > 0. Et *la solution continue non identiquement nulle la plus générale de* (F_1) *est représentable sous la forme*

$$D(s, t) = \frac{A(t)}{A(s)},$$

où $A(s)$ *est une fonction continue positive arbitrairement choisie.*

Ainsi nous venons de déterminer la solution continue la plus générale de (F_1), et même nous avons déterminé plus haut la solution — continue ou non — mais partout $\neq 0$, la plus générale. C'est ce qui nous suffira pour l'étude qui suit de l'équation plus générale (F_r). Il est intéressant cependant de montrer qu'on peut obtenir la solution la plus générale de (F_1). Mais ce résultat, moins utile pour les applications, va être reporté à la fin de ce Mémoire, sous le titre de *Note F* (p. 294).

Retour à l'équation fonctionnelle (F_r). — Revenons maintenant au cas général où $r \geqq 1$. Il est clair que si les $\varphi_{ik}(s, t)$ sont des fonctions continues du couple (s, t), il en sera de même de leur déterminant $D(s, t)$. Dès lors, celui-ci étant une solution continue de (F_1) ou bien est identiquement nul ou bien est partout > 0.

En résolvant, comme nous allons le faire, l'équation fonctionnelle (F_r) dans le cas général où $D(s, t)$ reste $\not\leq 0$, on aura donc effectué une résolution qui comprend comme cas très particulier le cas où l'on suppose que les $\varphi_{ik}(s, t)$ sont continues et que leur déterminant est $\not\leq 0$ pour au moins un couple (s, t). D'ailleurs, dans le cas où sont vérifiées les conditions

$$(L') \qquad \varphi_{jk}(s, s) = \begin{cases} 1 & \text{si } j = k, \\ 0 & \text{si } j \neq k, \end{cases}$$

analogues aux conditions (L) de l'application aux probabilités, on aura évidemment $D(s, s) = 1$.

Par conséquent, la solution continue la plus générale de l'ensemble des équations (F_r) et (L') est, en particulier, une solution continue de (F_r) telle que $D(s, t)$ reste > 0.

D'ailleurs, la résolution de (F_r) quand $D(s, t)$ reste $\not\leq 0$ donnera aussi des résultats utiles dans le cas général (étudié dans la note F) où l'on ne fait aucune hypothèse sur $D(s, t)$. En effet, s'il existe un couple (s_0, t_0) tel que $D(s_0, t_0) \not\leq 0$, alors il existe au moins un intervalle tel que $D(s, t)$ soit $\not\leq 0$ quand s, t sont intérieurs à cet intervalle, et tel, par conséquent, qu'on puisse appliquer à l'intérieur de cet intervalle la solution que nous allons déterminer.

Solution la plus générale de (F_r) quand le déterminant des $\Phi_{jk}(s, t)$ reste $\not\leq 0$. — Supposons que $D(s, t)$ soit $\not\leq 0$ pour $s \leq t$ ou au moins quand s, t sont intérieurs à un intervalle fixe (S, T) $(S < s \leq t < T)$.

Soit alors u_0 un point fixe intérieur à (S, T). Comme on l'a fait plus haut pour $D(s, t)$, on va, pour obtenir l'expression des $\varphi_{ik}(s, t)$, considérer trois cas :

I. $u_0 \leqq s \leqq t$; on a alors

$$\varphi_{ik}(u_0, t) = \sum_j \varphi_{ij}(u_0, s)\,\varphi_{jk}(s, t).$$

Pour chaque système de valeurs fixes de k, de s et de t, les $\varphi_{jk}(s, t)$ sont solutions des équations en X_j :

$$\varphi_{ik}(u_0, t) = \sum_{j=1}^{j=r} \varphi_{ij}(u_0, s) X_j \qquad (i = 1, \ldots, r).$$

Puisque $D(u_0, s) \neq 0$, ce système a un système unique de solutions donné par la règle de Cramer. On a donc

$$\varphi_{jk}(s, t) = \frac{1}{D(u_0, s)} \sum_i \Phi_{ij}(u_0, s)\, \varphi_{ik}(u_0, t),$$

où les $\Phi_{ij}(u_0, s)$ sont les coefficients des $\varphi_{ij}(u_0, s)$ dans le développement de $D(u_0, s)$

II. $s \leqq u_0 \leqq t$; on a dans ce cas

$$\varphi_{jk}(s, t) = \sum_i \varphi_{ji}(s, u_0)\, \varphi_{ik}(u_0, t)$$

III. $s \leqq t \leqq u_0$, et par suite

$$\varphi_{ji}(s, u_0) = \sum_k \varphi_{jk}(s, t)\, \varphi_{ki}(t, u_0) \qquad (i = 1, \ldots, r)$$

D'où, comme plus haut, puisque $D(t, u_0) \neq 0$,

$$\varphi_{jk}(s, t) = \frac{1}{D(t, u_0)} \sum_i \Phi_{ki}(t, u_0)\, \varphi_{ji}(s, u_0)$$

On voit qu'on peut alors mettre ces trois expressions de $\varphi_{jk}(s, t)$ sous la forme (commune à ces trois cas, c'est-à-dire valable si $S \leqq s \leqq t < T$)

$$(4) \qquad \varphi_{jk}(s, t) = \sum_i a_{ji}(s)\, b_{ki}(t)$$

Il suffit de poser

$$(5) \qquad a_{ji}(s) = \begin{cases} \varphi_{ji}(s, u_0) & \text{pour} \quad s \leqq u_0, \\ \dfrac{\Phi_{ij}(u_0, s)}{D(u_0, s)} & \text{pour} \quad u_0 \leqq s, \end{cases}$$

$$(6) \qquad b_{ki}(t) = \begin{cases} \dfrac{\Phi_{ki}(t, u_0)}{D(t, u_0)} & \text{pour} \quad t \leqq u_0, \\ \varphi_{ik}(u_0, t) & \text{pour} \quad u_0 \leqq t. \end{cases}$$

D'ailleurs, on ne peut choisir indépendamment les fonctions $a_{ji}(s)$,

$b_{ik}(t)$. On voit, en effet, qu'on a

$$\sum_i a_{ij}(s)\, b_{ik}(s) = \begin{cases} \dfrac{1}{D(s, u_0)} \displaystyle\sum_i \varphi_{ji}(s, u_0)\, \Phi_{ik}(s, u_0) & \text{si} \quad s \leq u_0, \\ \dfrac{1}{D(u_0, s)} \displaystyle\sum_i \Phi_{ji}(u_0, s)\, \varphi_{ki}(u_0, s) & \text{si} \quad u_0 \leq s \end{cases}$$

et, par suite, quel que soit s,

$$\sum_i a_{ij}(s)\, b_{ik}(s) = \begin{cases} 0 & \text{si} \quad j \neq k, \\ 1 & \text{si} \quad j = k. \end{cases} \tag{7}$$

Et de même

$$\sum_i a_{ji}(s)\, b_{ki}(s) = \begin{cases} 0 & \text{si} \quad j \neq k, \\ 1 & \text{si} \quad j = k. \end{cases}$$

Inversement, donnons-nous arbitrairement un système de fonctions $a_{ij}(s)$, $b_{kl}(s)$ biorthogonal et normé par rapport aux seconds indices, c'est-à-dire vérifiant les conditions (7). Et considérons les fonctions définies par

$$\varphi_{jk}(s, t) = \sum_l a_{jl}(s)\, b_{kl}(t). \tag{8}$$

On a

$$\begin{aligned} \sum_i \varphi_{ji}(s, u)\, \varphi_{ik}(u, t) &= \sum_i \Big[\sum_h a_{jh}(s)\, b_{ih}(u) \Big] \Big[\sum_l a_{il}(u)\, b_{kl}(t) \Big] \\ &= \sum_{h,l} \Big\{ a_{jh}(s)\, b_{kl}(t) \sum_i a_{il}(u)\, b_{ih}(u) \Big\} \\ &= \sum_l a_{jl}(s)\, b_{kl}(t) = \varphi_{jk}(s, t). \end{aligned}$$

D'autre part, si $D(s)$, $d(s)$ sont les déterminants des $b_{kl}(s)$ et des $a_{kl}(s)$, on voit que le produit $D(s)\, d(s)$ est un déterminant dont les termes sont nuls, sauf ceux de la diagonale principale, tous égaux à 1. On a donc

$$d(s)\, D(s) = 1.$$

Donc $d(s)$, $D(t)$ restent $\neq 0$. Or, $D(s, t) = d(s)\, D(t)$. Donc $D(s, t) \neq 0$.

En résumé, *la solution (continue ou non en s, t) $\varphi_{ik}(s, t)$ de (F_i), la plus générale parmi celles pour lesquelles le déterminant $D(s, t)$ des $\varphi_{ik}(s, t)$ reste $\neq 0$ est de la forme*

$$\varphi_{jk}(s, t) = \sum_i a_{ji}(s)\, b_{ki}(t),$$

où les $a_{ji}(s)$, $b_{ki}(s)$ sont un système biorthogonal et normé par rapport aux seconds indices, mais par ailleurs arbitraire.

Ce résultat s'applique, qu'on fasse varier les s, t arbitrairement sur la droite ou à l'intérieur d'un intervalle fixe (S, T).

Observons d'ailleurs qu'on peut choisir les $a_{ji}(s)$ presque arbitrairement, les $b_{ki}(s)$ étant alors déterminés. En effet, on a vu que $d(s) \neq 0$. D'autre part, l'ensemble des conditions exprimant que le système des a, b est biorthogonal et normé est équivalent au système (obtenu par simple résolution)

$$b_{ki}(s) = \frac{A_{ki}(s)}{d(s)},$$

où $A_{ki}(s)$ est le coefficient de $a_{ki}(s)$ dans le développement de $d(s)$.

En résumé, *la solution (continue ou non), $\varphi_{ik}(s, t)$, de l'équation fonctionnelle (F_i), la plus générale parmi celles pour lesquelles le déterminant des $\varphi_{ik}(s, t)$ reste $\neq 0$ pour $s \leqq t$, est de la forme*

$$\varphi_{jk}(s, t) = \sum_i a_{ji}(s)\, \frac{A_{ki}(t)}{d(t)}, \tag{9}$$

où les $a_{ji}(s)$ sont des fonctions qui peuvent être arbitrairement choisies pourvu que leur déterminant $d(s)$ reste $\neq 0$ et où $A_{ki}(t)$ est le coefficient de $a_{ki}(t)$ dans le déterminant $d(t)$.

Il est clair qu'on peut aussi écrire

$$\varphi_{jk}(s, t) = \sum_i \frac{B_{ji}(s)}{D(s)}\, b_{ki}(t), \tag{10}$$

où les $b_{ki}(t)$ sont choisies arbitrairement pourvu que leur déterminant $D(t)$ reste $\neq 0$, et où $B_{ji}(s)$ est le coefficient de $b_{ji}(s)$ dans le développement de $D(s)$.

On remarque d'ailleurs qu'on aura

$$\varphi_{jk}(s, s) = \frac{1}{d(s)} \sum_i a_{ji}(s) X_{ki}(s) = \begin{cases} 1 & \text{si } j = k, \\ 0 & \text{si } j \neq k. \end{cases}$$

De sorte que si $\varphi_{jk}(s, t)$ est une solution de (F_r) dont le déterminant $D(s, t)$ reste $\neq 0$, non seulement $D(s, s)$ reste égal à l'unité, mais encore les termes de ce déterminant des $\varphi_{jk}(s, s)$ sont nuls, sauf ceux de la diagonale principale qui sont égaux à 1, et par suite la condition (L′) est nécessairement vérifiée. On verra plus loin (p. 232) que la réciproque est vraie pour les solutions continues.

Remarque. — Tous nos raisonnements ont été encombrés par l'hypothèse que $s < t$ (laquelle s'impose d'elle-même dans l'application aux probabilités). Mais, à la négliger, nous risquions, *a priori*, d'écarter des solutions valables pour $s < t$ et non pour $s > t$. En fait, tant qu'il ne s'agit que de (F_r), ce risque n'existe pas. Car les fonctions (9) vérifient (F_r) *aussi bien pour $s > t$ que pour $s < t$*. On verra aussi, page 237 que si, en outre, elles vérifient (T′) pour $s < t$ elles satisfont également à (T′) pour $s > t$. Au contraire (et c'est là qu'apparaît l'importance de l'hypothese $s < t$), les formules (22) de la page, 242 relatives au cas de $r = 2$, montrent que des $\varphi_{jk}(s, t)$ vérifiant (F_r), (L′), (T′), (P′) pour $s < t$, ne vérifient pas nécessairement (P′) pour $s > t$.

Indétermination de la representation des $\varphi_{jk}(s, t)$. — Observons que pour une solution $\varphi_{jk}(s, t)$ de (F_r), il y a une infinité de manières de la mettre sous la forme

$$\varphi_{jk}(s, t) = \sum_i a_{ji}(s)\, b_{ki}(t) \tag{8}$$

Nous avons en effet déjà écrit, page 225, des formules (5), (6), donnant au moins un système d'expressions de $a_{ji}(s)$, $b_{ki}(t)$ pour $\varphi_{jk}(s, t)$ donné ; or, ces expressions dépendaient d'un paramètre arbitraire u_0.

Nous allons même obtenir *l'expression la plus générale des a_{ji}, b_{ki} pour une solution $\varphi_{jk}(s, t)$ donnée.*

Soit donc une expression

$$\varphi_{jk}(s, t) = \sum_i \alpha_{ji}(s)\, \beta_{ki}(t), \tag{11}$$

distincte ou non de la première. Nous commencerons par montrer que si c'est une solution de (F_i), et si le déterminant des $\varphi_{jk}(s, t)$ est $\neq 0$, le système des α, β est biorthogonal et normé (et les déterminants des α et des β sont $\neq 0$). Tout d'abord, le déterminant des $\varphi_{jk}(s, t)$ est, d'après l'expression (11), égal au produit des déterminants des α et des β. Le premier étant supposé $\neq 0$, il en sera de même des deux derniers. D'autre part, en écrivant que l'expression (11) vérifie (F_i), on aura

$$\sum_i \alpha_{ji}(s)\, \beta_{ki}(t) = \sum_{h,l} \alpha_{jh}(s)\, \beta_{kl}(t) \sum_i \beta_{ih}(u)\, \alpha_{li}(u)$$

$$= \sum_{h,l} G_{lh}(u)\, \alpha_{jh}(s)\, \beta_{kl}(t),$$

en posant

$$G_{lh}(u) = \sum_i \alpha_{li}(u)\, \beta_{ih}(u).$$

D'où

$$\sum_h \alpha_{jh}(s) \left[\beta_{kh}(t) - \sum_l G_{lh}(u)\, \beta_{kl}(t) \right] = 0.$$

Comme le déterminant des α est $\neq 0$, on voit que, pour chaque valeur de h, le crochet est nul. On a donc

$$\beta_{kh}(t) - \sum_l G_{lh}(u)\, \beta_{kl}(t) = 0,$$

et puisque le déterminant des $\beta_{ki}(t)$ est $\neq 0$, on a, pour chaque valeur fixe de h,

$$\sum_i \alpha_{li}(u)\, \beta_{ih}(u) = G_{lh}(u) = \begin{cases} 1 & \text{si } l = h, \\ 0 & \text{si } l \neq h, \end{cases}$$

comme il avait été annoncé.

Ceci étant, cherchons l'expression générale des α, β ; on a

$$\sum_k \alpha_{kh}(t)\, \varphi_{jk}(s, t) = \sum_i \alpha_{ji}(s) \left[\sum_k \alpha_{kh}(t)\, \beta_{ki}(t) \right] = \alpha_{jh}(s)$$

D'où

$$(12)\quad \left\{\begin{aligned} &\alpha_{jh}(s) = \sum_k \alpha_{kh}(t) \sum_l a_{jl}(s) b_{kl}(t) \\ &\qquad \text{c'est-à-dire} \\ &\alpha_{jh}(s) = \sum_l \gamma_{hl}\, a_{jl}(s) \end{aligned}\right.$$

où

$$\gamma_{hl} = \sum_k \alpha_{kh}(t)\, b_{kl}(t).$$

D'après cette expression, γ_{hl} pourrait dépendre de t, mais les équations (12) ont pour h fixe, $j = 1, \ldots, r$, un système unique de solutions γ_{hl}, puisque $d(s) \neq 0$ et ces solutions ne peuvent dépendre de t. Donc γ_{hl} est bien une constante. De façon analogue on aura

$$(13)\qquad \beta_{kl}(t) = \sum_i \delta_{li}\, b_{ki}(t) \qquad \text{où} \qquad \delta_{li} = \sum_j a_{ji}(s)\, \beta_{jl}(s) = \text{const}.$$

D'ailleurs les constantes γ_{hi} et δ_{li} ne sont pas indépendantes. Considérons en effet la somme

$$\sum_i \gamma_{hi}\delta_{li} = \sum_{k,j} \left[\sum_i a_{ji}(s)\, b_{ki}(s) \right] \alpha_{kh}(s)\, \beta_{jl}(s).$$

Les a, b étant nécessairement biorthogonaux et normés relativement aux seconds indices, le sont aussi par rapport aux premiers. On a donc

$$(14)\qquad \sum_i \gamma_{hi}\delta_{li} = \sum_j \alpha_{jh}(s)\, \beta_{jl}(s) = \begin{cases} 1 & \text{si } h = l, \\ 0 & \text{si } h \neq l, \end{cases}$$

et, par suite, aussi,

$$\sum_i \gamma_{ih}\delta_{il} = \begin{cases} 1 & \text{si } h = l, \\ 0 & \text{si } h \neq l. \end{cases}$$

Réciproquement, si les a_{ji}, b_{ki} sont donnés, et si l'on prend pour α_{jh}, β_{kl} des fonctions déterminées par

$$(12\ bis)\quad \left\{\begin{aligned} &\alpha_{jh}(s) = \sum_l \gamma_{hl}\, a_{jl}(s), \\ &\beta_{kl}(t) = \sum_i \delta_{li}\, b_{ki}(t), \end{aligned}\right.$$

en prenant pour les γ_{lh} et δ_{il} un système quelconque de constantes qui soit biorthogonal et normé par rapport au second indice, alors on aura aussi

$$\sum_l \alpha_{jl}(s)\beta_{kl}(t) = \sum_{h,l} a_{jh}(s)\,b_{kl}(t)\sum_l \gamma_{lh}\delta_{il} = \sum_l a_{jl}(s)\,b_{kl}(t) = \varphi_{jk}(s,t)$$

Remarques. — I. Soient $\delta(s)$, $\Delta(t)$ les déterminants des $\alpha_{jl}(s)$, $\beta_{kl}(t)$, on aura évidemment $D(s,t) = \delta(s)\Delta(t)$. Comme on a vu que dans le cas actuel où $D(s,t)$ est supposé toujours $\neq 0$, on a $D(s,s) = 1$, alors on voit que $\delta(s)\Delta(s) = 1$. Cela résulte aussi du fait que les α, β forment un système biorthogonal et normé. On en conclut :

$$D(s,t) = \frac{\delta(s)}{\delta(t)}.$$

II. Nous avons obtenu, d'une part, une représentation de la solution $\varphi_{jk}(s,t)$ de (F_r) sous la forme (4) en fonction d'un paramètre u_0, grâce aux formules (5), (6), d'autre part, la forme la plus générale d'une solution $\varphi_{jk}(s,t)$ donnée sous la forme (11). On pouvait se demander si la première méthode ne fournirait pas l'expression de $\varphi_{jk}(s,t)$ sous la forme (11) la plus générale quand on fait varier ce paramètre u_0. La réponse est négative. En effet, observons que si les $a_{jl}(s)$, $b_{kl}(t)$ sont obtenus par les formules (5) et (6), on a

$$a_{jl}(u_0) = b_{jl}(u_0) = \varphi_{jl}(u_0,u_0) = \begin{cases} 1 & \text{si } j = l, \\ 0 & \text{si } j \neq l \end{cases} \tag{15}$$

Or, puisqu'on peut choisir arbitrairement les fonctions $a_{jl}(s)$, pourvu que leur déterminant soit $\neq 0$, il est clair qu'on peut les choisir de sorte que les $a_{il}(s)$ ne satisfassent à la condition (15) pour aucune valeur de u_0. Il suffit, par exemple, de prendre pour les $a_{jl}(s)$ des constantes convenables.

Solution continue de (F_r) et (L'). — Considérons, en particulier, une solution $\varphi_{jk}(s,t)$ de (F_r) qui soit continue par rapport au couple (s,t) (pour $s \leq t$). Alors le déterminant $D(s,t)$ est une fonction continue; or, c'est une solution de (F_1), et nous avons vu (p. 222) qu'une solution continue de (F_1) est toujours positive ou toujours nulle.

I. Si la condition (L′) n'est pas vérifiée, c'est-à-dire s'il y a au moins une valeur de u et un couple j, k tel que $\varphi_{jk}(u, u) \neq 0$ si $j \neq k$, ou tel que $\varphi_{jj}(u, u) \neq 1$ si $j = k$, alors, comme on l'a vu (p. 228), $D(s, t)$ ne peut rester > 0 et, par suite, $D(s, t) \equiv 0$.

Ainsi, *pour toute solution continue de* (F_r) *ne vérifiant pas* (L′), *il existe un système de fonctions continues non toutes nulles* $U_k(s, t)$ *telles que*

$$\sum_k U_k(s, t)\,\varphi_{jk}(s, t) = 0.$$

II. Si, au contraire, $\varphi_{jk}(s, t)$ est une fonction continue vérifiant à la fois (F_r) et (L′), $D(s, s)$ reste égal à 1, donc $D(s, t)$ ne pouvant être identiquement nul reste positif.

En particulier, $D(s, t)$ restant > 0, $\varphi_{jk}(s, t)$ peut se mettre sous la forme (4) au moyen des formules (5), (6). D'après ces dernières formules, les $a_{ik}(s)$, $b_{ik}(t)$ sont des fonctions continues de s et de t. Mettons maintenant $\varphi_{jk}(s, t)$ sous la forme générale

$$\varphi_{jk}(s, t) = \sum_i \alpha_{ji}(s)\,\beta_{ki}(t). \tag{11}$$

Nous avons vu (p. 230) que les $\alpha_{ji}(s)$ sont nécessairement les combinaisons linéaires des $a_{ji}(s)$; les $\alpha_{ji}(s)$ seront donc continues, de même, pour les $\beta_{ki}(t)$. Réciproquement, si les $\alpha_{ji}(s)$ et les $\beta_{ki}(t)$ sont continues, l'expression (11) est aussi continue.

Ainsi, *la solution continue la plus générale des équations simultanées* (F_r) *et* (L′) *est représentable sous la forme* (11) *où les* $\alpha_{ji}(s)$, $\beta_{ki}(t)$ *sont des fonctions continues formant un système biorthogonal et normé arbitrairement choisi*. De plus, pour une solution continue $\varphi_{jk}(s, t)$ de (F_r) et de (L′), *chacune* de ses représentations sous la forme (11) est nécessairement constituée par un système biorthogonal et normé de fonctions continues.

On voit aussi que *la solution continue la plus générale des équations simultanées* (F_r) *et* (L′) *est de la forme*

$$\varphi_{jk}(s, t) = \sum_i a_{ji}(s)\,\frac{A_{ki}(t)}{d(t)},$$

où $A_{ki}(t)$ *est le coefficient de* $a_{ki}(t)$ *dans le développement du*

déterminant $d(t)$ *des* $a_{ki}(t)$ *et où les* $a_{ji}(s)$ *sont des fonctions continues dont le choix est arbitraire pourvu que leur déterminant* $d(s)$ *reste* $\neq 0$.

Notons que si c'est seulement à l'intérieur d'un intervervalle (S, T) — fini ou non — que $\varphi_{jk}(s, t)$ est solution continue simultanée de (F_{i}) et de (L'), alors l'énoncé précédent subsiste quand on ne l'applique qu'à l'*intérieur* de l'intervalle (S, T).

Comportement asymptotique des solutions continues. — En vue des applications à la théorie des probabilités en chaîne, nous allons étudier ce que deviennent les solutions continues $\varphi_{jk}(s, t)$ du système (F_{i}), (L') quand, s restant fixe, t croît indéfiniment, suivant ce que sont les fonctions continues arbitraires $b_{ki}(t)$ dans l'expression

$$(16) \qquad \varphi_{jk}(s, t) = \sum_{i} \frac{B_{ji}(s)\, b_{ki}(t)}{D(s)},$$

où $D(s)$ est le déterminant des $b_{ki}(s)$.

Toute cette étude se fera facilement si l'on se base sur ce que, comme on l'a vu pages 228 et 230 (en permutant les notations employées plus haut), les $b_{ki}(t)$ sont des combinaisons linéaires *à coefficients constants* des $\varphi_{ji}(u_0, t)$, soit

$$(17) \qquad b_{ki}(t) = \sum_{j} \delta_{ij}\, \varphi_{jk}(u_0, t),$$

dès que t devient supérieur à un nombre u_0 fixé arbitrairement d'avance

1° Pour qu'il existe au moins une valeur de s telle que les $\varphi_{ji}(s, t)$ restent bornés pour $t \geqq s$, il faut, d'après (17), et il suffit, d'après (16), que les $b_{ki}(t)$ soient bornés pour t assez grand. Et alors les $\varphi_{jk}(s, t)$ resteront bornés pour chaque valeur fixe de s quand t croît.

2° Pour qu'il existe au moins une valeur de s, telle que les $\varphi_{jk}(s, t)$ tendent chacun vers une limite déterminée et finie quand t croît, s restant fixe, il faut, d'apres (17), et il suffit, d'après (16), que les $b_{ki}(t)$ convergent quand t croît indéfiniment. Et alors, il en sera de même pour $\varphi_{jk}(s, t)$ pour toute valeur de s. Et si l'on a

$$\lim_{t \to +\infty} b_{ki}(t) = b_{ki},$$

on aura

$$\lim_{t \to \infty} \varphi_{jk}(s, t) = \sum_i \frac{B_{ji}(s)}{D(s)} b_{ki} = \varphi_{jk}(s).$$

D'ailleurs, on peut présenter les choses plus simplement, il suffit de choisir pour les mêmes solutions $\varphi_{jk}(s, t)$ une autre représentation

$$\varphi_{jk}(s, t) = \sum_i \alpha_{ji}(s) \beta_{ki}(t).$$

Si la relation entre les b et les β est de la forme

$$b_{ki}(t) = \sum_j \lambda_{ij} \beta_{kj}(t),$$

et si les $\beta_{kj}(t)$ tendent vers des limites β_{kj}, alors les $b_{ki}(t)$ tendent vers des limites b_{ki}, et l'on a

$$b_{ki} = \sum_j \lambda_{ij} \beta_{kj}$$

On sait que le déterminant des λ_{kj} est $\gtrless 0$. Donc le déterminant de b_{ki} et celui de β_{kj} sont en même temps nuls ou en même temps $\gtrless 0$.

Si le déterminant des b_{ki} est $\gtrless 0$, alors on pourra choisir les λ_{ij}, de sorte qu'on ait

$$\beta_{kj} = \begin{cases} 0 & \text{si } k \neq j, \\ 1 & \text{si } k = j. \end{cases}$$

Il suffira de prendre $\lambda_{ik} = b_{ki}$. Ainsi, *quand* le déterminant des limites de $\varphi_{kj}(s, t)$ est $\gtrless 0$, on pourra toujours poser [si les $\varphi_{jk}(s, t)$ convergent au moins pour une valeur de s quand t croît]

$$\varphi_{jk}(s, t) = \sum_i \alpha_{ji}(s) \beta_{ki}(t)$$

avec

$$\lim_{t \to \infty} \beta_{kj}(t) = \begin{cases} 0 & \text{si } k \neq j, \\ 1 & \text{si } k = j. \end{cases}$$

On aura alors

$$\alpha_{jk}(s) = \lim_{t \to \infty} \varphi_{jk}(s, t).$$

3° Pour qu'il existe au moins une valeur de s, telle que les $\varphi_{jk}(s, t)$ convergent chacun vers une limite $\varphi_k(s)$ déterminée, finie et indépen-

dante du premier indice. j, il faut, d'après (17), que les $b_{ki}(t)$ tendent vers des limites b_{ki}, finies, déterminées et de la forme de produits $\alpha_k c_i$ d'une fonction de k par une fonction de i. Et alors, on aura, d'après (16),

$$\varphi_k(s) = \alpha_k \sum_i \frac{B_{ji}(s) c_i}{D(s)}.$$

Il faudra donc encore : *a.* ou bien que les limites des $b_{ki}(t)$ soient toutes nulles [et alors les limites des $\varphi_{jk}(s, t)$ sont toutes nulles, et inversement]; *b.* ou bien que l'expression $\sum_i c_i B_{ji}(s)$ soit indépendante de j.

L'ensemble des conditions nécessaires qu'on vient d'établir est évidemment suffisant.

On notera que, dans ce cas, les lignes du déterminant limite des $b_{ik}(t)$ deviennent proportionnelles, donc $D(t)$ tend vers zéro. Et comme les $\varphi_{jk}(s, t)$ ont des limites finies, $D(s, t) = d(s) D(t)$ a une limite finie, et $d(t) = \frac{1}{D(t)}$ tend vers l'infini avec t.

4° Pour qu'il existe au moins une valeur de s, telle que les $\varphi_{jk}(s, t)$ soient des fonctions périodiques de t (de même période T), il faut, d'après (17), et il suffit, d'après (16), que les $b_{ki}(t)$ soient des fonctions périodiques de t (de période T). Et alors il en sera de même pour les $\varphi_{jk}(s, t)$, pour toute valeur de s, et la période sera indépendante de s.

5° Pour qu'il existe au moins une valeur de s telle que les fonctions $\varphi_{jk}(s, t)$ soient des fonctions asymptotiquement périodiques de t (et de même période asymptotique T), c'est-à-dire pour que $\varphi_{jk}(s, t)$ soit la somme d'une fonction de s et de t périodique en t (de période T indépendante de j, k) et d'une fonction de s et t qui converge vers zéro quand t croît — il faut, d'après (17), et il suffit, d'après (16), que les fonctions $b_{ki}(t)$ soient des fonctions asymptotiquement périodiques et de même période T. Et alors, il en sera de même, quel que soit s, et la période asymptotique sera indépendante de s.

Limites généralisées. — On a généralisé de diverses façons la limite d'une fonction de t lorsque t croît indéfiniment. Nous ne

Ainsi, il faut que les quantités $c_i(t)$ se réduisent à des nombres c_i indépendants de t, et que l'on ait

$$1 = \sum_i c_i a_{ji}(s).$$

Réciproquement, supposons les $a_{ji}(s)$ arbitraires, sauf : 1° que leur déterminant $d(s)$ soit $\neq 0$, 2° qu'il existe des nombres indépendants de s, c_i, tels que

$$(18\ bis) \qquad \sum_i c_i a_{ji}(s) = 1 \qquad (j = 1, \ldots, r)$$

Alors on tire de ces équations, comme plus haut,

$$\sum_k b_{ki}(s) = c_i$$

et, par suite,

$$1 = \sum_i \left[\sum_k b_{ki}(t) \right] a_{ji}(s) = \sum_k \sum_i a_{ji}(s) b_{ki}(t) = \sum_k \varphi_{jk}(s, t)$$

L'ensemble des conditions 1°, 2° est donc nécessaire et suffisant pour assurer la condition

$$(\mathrm{T}') \qquad \sum_k \varphi_{ik}(s, t) = 1 \qquad (i = 1, \ldots, r)$$

Il est clair que $c_1, \ldots, c_r$ ne peuvent être tous nuls. Si, par exemple, $c_r \neq 0$, on voit que pour assurer la condition 2°, il suffira de prendre les $a_{jr}(s)$ tels que

$$a_{jr}(s) = \frac{1 - \sum_{i=1}^{i=r-1} c_i a_{ji}(s)}{c_r} \qquad (j = 1, \ldots, r)$$

On peut aussi assurer la condition (T′) d'une autre façon, en profitant de l'indétermination du système des a, b. On a vu qu'on pouvait lui substituer un système α, β au moyen des formules (12), (13). Si (T′) est vérifié, il devra y avoir un système de constantes c_i,

c'_h telles que

$$\sum_i c_i a_{ji}(s) = 1 = \sum_h c'_h \alpha_{jh}(s) = \sum_i a_{ji}(s) \sum_h c'_h \gamma_{hi}.$$

Comme $\sum_i c_i a_{ji}(s) = 1$, il suffit de prendre les c'_h tels que

$$c_i = \sum_h c'_h \gamma_{hi}$$

Choisissons les c'_h.

Pour qu'on ait $c'_1 = 1, c'_2 = \ldots = c'_l = 0$, il suffit de prendre $\gamma_{1i} = c_i$. Si donc (T') est vérifiée, on pourra, — en prenant les $\gamma_{1i} = c_i$ (les c_i ne peuvent être tous nuls) —, puis les γ_{ji} ($j \neq 1$), de sorte que le déterminant des γ soit $\neq 0$, puis en déterminant les δ_{ki} au moyen des γ et enfin les α, β au moyen des a, b, γ, δ, — mettre φ_{jk} sous la forme (11) avec $\alpha_{j1}(s) = 1$. D'où

$$\varphi_{jk}(s, t) = \beta_{1k}(t) + \sum_{i=2}^{i=l} \alpha_{ji}(s)\, \beta_{ik}(t)$$

D'ailleurs, le système des α, β étant biorthogonal et normé par rapport aux seconds indices, on aura

$$\sum_j \alpha_{j1}(s)\, \beta_{jh}(s) = \begin{cases} 1 & \text{si } h = 1, \\ 0 & \text{si } h \neq 1. \end{cases}$$

D'où

$$\sum_j \beta_{j1}(s) = 1 \qquad \text{et} \qquad \sum_j \beta_{jh}(s) = 0 \qquad \text{pour} \quad h \neq 1$$

Par suite

$$\sum_k \varphi_{jk}(s, t) = \sum_i \left[\alpha_{ji}(s) \sum_k \beta_{ki}(t) \right] = 1.$$

En résumé, *pour qu'une solution des équations simultanées* (F_r) *et* (L') *vérifie la condition* (T'), $\sum_k \varphi_{jk}(s, t) = 1$, *il faut et il suffit que l'on puisse choisir l'une des manières de la mettre sous la forme* (11), *de façon que les* $\alpha_{j1}(s)$ *soient tous identiques à* 1, les α et β continuant à former un système biorthogonal et normé [et les α, β seront continues si la solution $\varphi_{jk}(s, t)$ est continue].

Le comportement asymptotique des solutions continues du sys-

tème (F_t), (L') donne lieu à quelques remarques quand la condition (T') est vérifiée par ces solutions.

Tout d'abord si les $\varphi_{jk}(s, t)$ ont des limites $\varphi_{jk}(s)$ lorsque t croît indéfiniment, ces limites vérifieront évidemment aussi la condition, limite de (T'),

$$\sum_k \varphi_{jk}(s) = 1, \tag{19}$$

remarque évidente mais parfois très importante.

On voit aussi qu'on aura

$$\varphi_{jk}(s) = \sum_i \varphi_{ji}(s, u)\,\varphi_{ik}(u) \tag{20}$$

Si, en outre, les $\varphi_{jk}(s)$ sont des quantités $\varphi_k(s)$ indépendantes du premier indice, l'égalité précédente devient

$$\varphi_k(s) = \varphi_k(u)\sum_i \varphi_{ji}(s, u) = \varphi_k(u)$$

Ainsi *lorsqu'une solution continue* $\varphi_{jk}(s, t)$ *du système* (F_t), (L'), (T') *converge quand* t *croît indéfiniment vers une limite indépendante du premier indice,* j, *cette limite est aussi indépendante de la première variable,* s.

D'ailleurs, *réciproquement*, si une solution continue $\varphi_{jk}(s, t)$ du système (F_t), (L') converge quand t croît indéfiniment vers une limite φ_k indépendante à la fois du premier indice j et de la première variable s, alors *ou bien cette limite est nulle quel que soit* k, *ou bien la solution* $\varphi_{jk}(s, t)$ *vérifie la condition* (T'). Car l'égalité (20) devient

$$\varphi_k\left[1 - \sum_i \varphi_{ji}(s, u)\right] = 0$$

Remarques sur les solutions non négatives. — Lorsque la solution $\varphi_{jk}(s, t)$ de (F_t) vérifie à la fois les conditions (T') et

$$\varphi_{jk}(s, t) \geqq 0, \tag{P'}$$

on a aussi nécessairement $\varphi_{jk}(s, t) \leqq 1$. On en déduit [p. 233, formule (17)] que, dans ce cas, si $D(s, t)$ reste $\neq 0$, alors quelle que soit la représentation (11) de $\varphi_{jk}(s, t)$, les $\beta_{ki}(t)$ restent bornés quand

$t \to +\infty$. [Puisque dans (17), $t > u_0$, le raisonnement ne renseigne pas sur le comportement de $\beta_{ki}(t)$ quand $t \to -\infty$.]

De la même manière, on montre que les $\alpha_{ji}(s)$ sont bornés quand $s \to -\infty$.

On a, en revenant aux notations a, b,

$$a_{ji}(s) = \frac{B_{ji}(s)}{D(s)} = B_{ji}(s)\,d(s), \tag{21}$$

ce qui va permettre d'étudier le cas où $s \to +\infty$ les déterminants $B_{ji}(s)$ formés avec des termes $b_{ji}(s)$ qui restent bornés, quand $s \to +\infty$ restent donc aussi bornés. Pour que l'un au moins des $a_{ji}(s)$ ne soit pas borné, il suffit donc que $d(s)$ ne le soit pas. Inversement, il va de soi que si les termes $a_{ji}(s)$ sont bornés, leur déterminant $d(s)$ sera borné. Ainsi : la condition nécessaire et suffisante pour que l'un au moins de $a_{ji}(s)$ ne soit pas borné quand $s \to +\infty$ est que leur déterminant $d(s)$ ne le soit pas non plus.

On voit l'importance du comportement de $d(s)$ quand $s \to +\infty$. Or, on peut préciser celui-ci quand les conditions (P'), (T') sont vérifiées. Commençons par examiner le déterminant $D(s, t)$ des $\varphi_{jk}(s, t)$.

D'une façon générale, si les termes u_{jk} d'un déterminant vérifient les conditions

$$u_{jk} \geqq 0, \qquad \sum_k u_{jk} \leqq 1,$$

la valeur absolue de ce déterminant est $\leqq 1$. En effet, les mêmes conditions seront vérifiées par les mineurs de ce déterminant. Si donc la propriété énoncée, évidemment vraie pour un déterminant d'ordre 2, est vraie pour des déterminants d'ordre $r-1$, on aura, en développant un déterminant Δ d'ordre r, une expression de la forme

$$\Delta = \sum_k u_{jk}\,U_{jk},$$

et le déterminant U_{jk}, d'ordre $r-1$, sera en valeur absolue $\leqq 1$. D'où

$$|\Delta| \leqq \sum_k |u_{jk}| \leqq 1.$$

Il en résulte que pour toute solution $\varphi_{jk}(s, t)$ de (F_r) vérifiant les

conditions (P'), (T'), on aura

$$|D(s, t)| \leq 1$$

Si, de plus, pour cette solution $D(s, t)$ reste $\neq 0$, on aura

$$1 = D(s, s) = d(s) D(s),$$

d'où

$$(21\ bis) \qquad D(s, t) = \frac{d(s)}{d(t)},$$

et enfin

$$\left|\frac{d(s)}{d(t)}\right| \leq 1$$

Ainsi, pour toute solution $\varphi_{jk}(s, t)$ de (F_1) vérifiant les conditions (P'), (T') et dont le déterminant $D(s, t)$ reste $\neq 0$, la fonction $|d(s)|$ est positive et non décroissante.

Si, en outre, les $\varphi_{jk}(s, t)$ sont continues, $D(s, t)$ reste $\neq 0$, et $d(s)$ a un signe constant. En changeant au besoin de signe les $a_{i2}(s)$, par exemple [de sorte qu'on peut laisser les $a_{i1}(s) = 1$], et en changeant en conséquence les $b_{i2}(t) = \frac{\Delta_{i2}(t)}{d(t)}$, les $\varphi_{jk}(s, t)$ ne sont pas changés et $d(s)$ étant ainsi changé de signe pourra être supposé > 0.

En résumé, *si $\varphi_{jk}(s, t)$ est une solution continue de (F_1) qui vérifie les conditions (L'), (P'), (T'), elle peut être mise sous la forme (9) où les $a_{ji}(s)$ sont des fonctions continues, bornées quand $s \to -\infty$, où les $a_{j1}(s) \equiv 1$, où le déterminant $d(s)$ des $a_{ji}(s)$ est une fonction positive non décroissante de s, et où les termes $b_{ji}(t) = \frac{\Delta_{ji}(t)}{d(t)}$ sont bornés quant $t \to +\infty$.* Et, pour que les $a_{ji}(s)$ et $b_{ji}(s)$ soient bornés quel que soit s, il faut et il suffit que le déterminant $d(s)$ soit borné quand s tend vers $+\infty$.

Solution continue la plus générale de (F_1) vérifiant les conditions (L'), (T'), (P') dans le cas où $r = 2$. — Par exemple, si $r = 2$, on pourra toujours poser en vertu de (T')

$$\alpha_{11}(s) = 1, \qquad \alpha_{21}(s) = 1, \qquad \alpha_{12}(s) = -a(s), \qquad \alpha_{22}(s) = b(s);$$

on aura, en vertu des conditions d'orthogonalité des α, β,

$$d(s) = b(s) + a(s),$$

$$\beta_{11}(t) = 1 - \frac{a(t)}{d(t)}, \qquad \beta_{12}(t) = \frac{1}{d(t)}, \qquad \beta_{21} = \frac{a(t)}{d(t)}, \qquad \beta_{22}(t) = -\frac{1}{d(t)},$$

d'où

$$(22) \quad \begin{cases} \varphi_{11}(s, t) = 1 - \dfrac{a(t) - a(s)}{d(t)}, & \varphi_{21}(s, t) = \dfrac{b(t) - b(s)}{d(t)}, \\ \varphi_{12}(s, t) = \dfrac{a(t) - a(s)}{d(t)}, & \varphi_{22}(s, t) = 1 - \dfrac{b(t) - b(s)}{d(t)}. \end{cases}$$

Pour $r = 2$, si les $\varphi_{jk}(s, t)$ sont continues et si la condition (L′) et la condition (T′) sont remplies, les φ_{jk} sont de cette forme avec $d(t)$ constamment $\neq 0$, donc $d(t)$ d'un signe constant. On peut supposer $d(s) > 0$, car si $d(s)$ était < 0, il suffirait de poser $a_1(s) = -a(s)$, $b_1(s) = -b(s)$, et alors $d_1(s) = -d(s)$ serait > 0.

Si l'on veut, de plus, que les $\varphi_{jk}(s, t)$ soit tous $\geqq 0$ et $\leqq 1$, il suffit, en vertu de (T′), d'écrire qu'ils sont $\geqq 0$. Il faut d'abord $\varphi_{12} \geqq 0$ et $\varphi_{21} \geqq 0$ donc que $a(t) \geqq a(s)$ et $b(t) \geqq b(s)$ pour $s < t$: autrement dit, $a(s)$ et $b(s)$ sont non décroissantes.

Les conditions nécessaires déjà formulées sont alors suffisantes. En effet, pour $\varphi_{11} \geqq 0$, il suffit que $[b(t) + a(s)] \geqq 0$ et pour $\varphi_{22} \geqq 0$, que $[b(s) + a(t)] \geqq 0$.

Or,

$$b(t) + a(s) = [b(t) - b(s)] + d(s)$$

et

$$b(s) + a(t) = d(s) + [a(t) - a(s)],$$

et ces deux seconds membres sont > 0 quand on suppose $d(s)$ toujours > 0 et les deux fonctions $a(s)$, $b(s)$ non décroissantes.

Ainsi les solutions continues les plus générales de l'application aux probabilités en chaîne sont données dans le cas de $r = 2$ par les formules (22) où $d(s) = b(s) + a(s)$ et où $a(s)$, $b(s)$ sont des fonctions continues non décroissantes dont la somme reste positive.

Il est d'ailleurs facile dans ce cas de déterminer l'allure asymptotique de $\varphi_{jk}(s, t)$ lorsque $t \to +\infty$.

En effet, les fonctions $a(t)$, $b(t)$, $d(t)$ étant non décroissantes, tendent chacune vers une limite finie ou vers $+\infty$.

Pour que $d(t) = a(t) + b(t)$ ait une limite finie C quand $t \to +\infty$, il faut et il suffit que $a(t)$ et $b(t)$ tendent respectivement vers des limites finies que nous appellerons A et B.

Si $d(t) \to C$, alors $a(t) \to A$, $b(t) \to B$, et l'on voit immédiatement que l'on a

$$\lim_{t \to +\infty} \varphi_{11}(s, t) = \frac{B + a(s)}{B + A}, \qquad \lim_{t \to +\infty} \varphi_{12}(s, t) = \frac{A - a(s)}{B + A},$$

$$\lim_{t \to +\infty} \varphi_{21}(s, t) = \frac{B - b(s)}{B + A}, \qquad \lim_{t \to +\infty} \varphi_{22}(s, t) = \frac{A + b(s)}{A + B}.$$

La limite de $\varphi_{jk}(s, t)$ n'est d'ailleurs jamais indépendante du premier indice dans le cas actuel.

Si $d(t) \to +\infty$, on voit que, dans tous les cas, les quatre sommes

$$\varphi_{11}(s, t) + \varphi_{22}(s, t), \quad \varphi_{21}(s, t) + \varphi_{22}(s, t),$$

$$\varphi_{21}(s, t) + \varphi_{12}(s, t), \quad \varphi_{11}(s, t) + \varphi_{22}(s, t)$$

tendent vers l'unité, car

$$\varphi_{11}(s, t) + \varphi_{12}(s, t) = 1 = \varphi_{21}(s, t) + \varphi_{22}(s, t)$$

$$\varphi_{21}(s, t) + \varphi_{12}(s, t) = 1 - \frac{d(s)}{d(t)},$$

$$\varphi_{11}(s, t) + \varphi_{22}(s, t) = 1 + \frac{d(s)}{d(t)}.$$

Mais il peut se trouver que $a(t)$ ou $b(t)$ ait une limite finie, on a donc les cas suivants :

$$a(t) \to A, \quad b(t) \to +\infty, \quad a(t) \to +\infty, \quad b(t) \to B, \quad a(t) \to +\infty, \quad b(t) \to +\infty$$

Dans les deux premiers cas, les $\varphi_{jk}(s, t)$ ont encore des limites déterminées :

$$\left.\begin{matrix} 1 & 0 \\ 1 & 0 \end{matrix}\right\} \quad [si\ b(t) \to +\infty];$$

$$\left.\begin{matrix} 0 & 1 \\ 0 & 1 \end{matrix}\right\} \quad [si\ a(t) \to +\infty].$$

La limite de $\varphi_{jk}(s, t)$ est dans le cas actuel indépendante de j et de s.

Enfin, si $a(t)$ et $b(t)$ tendent simultanément vers l'infini, et

si $\varphi_{21}(s, t)$ et $\varphi_{12}(s, t)$ ont chacun une limite, les égalités

$$\frac{b(t)}{d(t)} = \varphi_{21}(s, t) + \frac{b(s)}{d(t)},$$

$$\frac{a(t)}{d(t)} = \varphi_{11}(s, t) + \frac{a(s)}{d(t)}$$

montrent que $\frac{b(t)}{d(t)}$ et $\frac{a(t)}{d(t)}$ tendent vers des limites respectives β et α (avec $\alpha + \beta = 1$). Et réciproquement, cela suffit pour que les $\varphi_{jk}(s, t)$ convergent, avec

$$\lim_{t \to +\infty} \varphi_{11}(s, t) = \beta = \lim_{t \to +\infty} \varphi_{21}(s, t),$$

$$\lim_{t \to +\infty} \varphi_{12}(s, t) = \alpha = \lim_{t \to +\infty} \varphi_{22}(s, t).$$

La limite de $\varphi_{jk}(s, t)$ est donc encore dans ce cas indépendante de j et de s.

On peut résumer ce qui précède en disant :

Pour que les quatre fonctions $\varphi_{jk}(s, t)$ convergent quand t croît indéfiniment, il faut et il suffit que la courbe plane $x = a(t)$, $y = b(t)$ soit bornée quand t croît ou que sa branche infinie, quant t croît, possède une direction asymptotique déterminée.

Pour que, dans ce cas, la limite de $\varphi_{jk}(s, t)$ soit indépendante de j (et alors nécessairement indépendante de s), il faut et il suffit que $d(t)$ tende vers $+\infty$ avec t.

Observons qu'on a, dans tous les cas,

$$\varphi_{11}(s, t) - \varphi_{21}(s, t) = \frac{d(s)}{d(t)} = \varphi_{22}(s, t) - \varphi_{12}(s, t).$$

Donc

$$\varphi_{11}(s, t) - \varphi_{21}(s, t) \quad \text{et} \quad \varphi_{22}(s, t) - \varphi_{12}(s, t)$$

convergent toujours vers une limite, qui est nulle si $d(t) \to +\infty$ et qui est $\neq 0$ $\left(\text{et égale à } \frac{d(s)}{C} > 0\right)$ dans le cas contraire.

Donc, *si l'on attribue à chacune des quatre fonctions $\varphi_{jk}(s, t)$ une limite généralisée (satisfaisant aux conditions énoncées* p. 236), *la condition nécessaire et suffisante pour que, quand t croît, la limite généralisée de $\varphi_{jk}(s, t)$ soit indépendante de j (et alors indépendante de s) est que $d(t)$ tende vers $+\infty$ avec t.*

Cas de $r > 2$. — Nous allons montrer comment on peut traiter le cas où r est un entier quelconque. Tout d'abord les résultats obtenus par la quatrième méthode, concernant la solution la plus générale du système (I), (T′), (L′), peuvent être formulés ainsi : la solution continue la plus générale, pour r entier quelconque, du système des conditions (I), (T′), (L′) et (P′), est fournie par les formules

$$(25) \qquad P_{jk}(s, t) = \sum_{l} \frac{a_{jl}(s)\, A_{kl}(t)}{d(t)},$$

où $A_{kl}(t)$ est le coefficient de $a_{kl}(t)$ dans le développement du déterminant $d(t)$ des $a_{jl}(t)$, et où les $a_{jl}(s)$ sont des fonctions continues de s prises arbitrairement parmi celles qui satisfont aux conditions suivantes : leur déterminant $d(t)$ est constamment positif, les $a_{j1}(s)$ sont égaux à 1, les sommes $\sum_{l} a_{jl}(s)\, A_{kl}(t)$ sont ≥ 0 pour $s \leq t$.

En utilisant un théorème de Kolmogoroff, nous allons simplifier la vérification de la dernière condition de façon à ne considérer les signes que de fonctions *d'une seule variable* (Fréchet [151]) *quand on se limite aux solutions dérivables*, à dérivées continues.

Pour cela, considérons les solutions continues $P_{jk}(s, t)$ qui sont dérivables pour $s \leq t$.

Nous allons d'abord montrer qu'elles sont alors nécessairement dérivables en s et t pour tout système de valeurs s_0, t_0 de s et de t. En effet, nous avons montré (p. 230) que dans toute représentation de la forme

$$P_{jk}(s, t) = \sum_{l} \alpha_{jl}(s)\, \beta_{kl}(t),$$

les $\alpha_{jl}(s)$ sont des combinaisons linéaires des fonctions $P_{jl}(s, u_0)$ pour $s < u_0$, et comme on peut prendre u_0 arbitraire et en particulier $> s_0$, on voit que $P_{jl}(s, u_0)$ et, par suite $\alpha_{jl}(s)$, sera dérivable en s pour $s \leq u_0$, et en particulier pour $s = s_0$. On verrait de même que $\beta_{kl}(t)$ est dérivable pour $t = t_0$. Ainsi $\alpha_{jl}(s)$ et $\beta_{jl}(t)$ sont dérivables en s et t pour toute valeur s_0 de s, t_0 de t et par suite $P_{jk}(s, t)$ est aussi dérivable quels que soient s et t.

Ceci étant, on voit, en posant

$$U_{jk}(t) = \lim_{\delta > 0} \frac{P_{jk}(t, t+\delta) - P_{jk}(t, t)}{\delta},$$

que l'on a

$$U_{jk}(t) = \sum_i \alpha_{ji}(t)\,\beta'_{ki}(t)$$

et puisque

$$(24) \qquad \sum_i \alpha_{ji}(t)\,\beta_{ki}(t) = \begin{cases} 0 & \text{si } j \neq k, \\ 1 & \text{si } j = k, \end{cases}$$

$$U_{jk}(t) = -\sum_i \alpha'_{ji}(t)\,\beta_{ki}(t).$$

De plus, en vertu des conditions et des identités $\alpha_{j1}(t) = 1$ et des identités

$$\sum_j \alpha_{ji}(t)\,\beta_{jk}(t) = \begin{cases} 0 & \text{si } i \neq k, \\ 1 & \text{si } i = k, \end{cases}$$

déduites de (24), on a

$$\sum_j \beta_{jk}(t) = \begin{cases} 0 & \text{si } k \neq 1, \\ 1 & \text{si } k = 1, \end{cases}$$

et, par suite,

$$(25) \qquad \sum_k U_{jk}(t) = 0.$$

Or, il a été démontré plus haut que l'on a

$$(26) \qquad U_{jk}(t) \leq 0 \quad \text{pour } j \neq k, \qquad U_{kk}(t) \geq 0.$$

On a, d'autre part, en revenant aux notations a, b,

$$U_{jk}(t) = -\sum_i a'_{ji}(t)\,\frac{A_{ki}(t)}{d(t)}.$$

Ainsi les $a_{ji}(t)$ sont des fonctions continues dérivables et telles [puisque $d(t) > 0$] que l'on ait

$$(27) \qquad \sum_i a'_{ji}(t)\,A_{ki}(t) \begin{cases} \leq 0 & \text{pour } j \neq k, \\ \geq 0 & \text{pour } j = k. \end{cases}$$

Réciproquement, supposons qu'en outre des conditions mentionnées plus haut, ces dernières conditions soient remplies. Alors l'expression (23) sera une solution continue et dérivable du système

des conditions (I'), (L'), (T'). Elle vérifiera le système

$$\frac{\partial P_{jk}(s, t)}{\partial t} = \sum_l U_{jk}(t) P_{lj}(s, t),$$

et les $U_{jk}(t)$ vérifieront (25), et, en vertu de (27), les conditions (26). En outre, ce sont les solutions de ce système qui pour $t = s$ prennent en vertu de (L') les valeurs

$$P_{jk}(t, t) = \begin{cases} 0 & \text{si} \quad j \neq k, \\ 1 & \text{si} \quad j = k \end{cases}$$

Or, il a été prouvé plus haut que de telles solutions sont nécessairement ≥ 0. La solution considérée vérifie donc aussi la condition (P').

En résumé, *la solution continue la plus générale vérifiant le système* (I), (P), (L), (T) *parmi celles qui sont dérivables en s et t pour $s < t$: 1° Est aussi dérivable aux points où $s = t$; 2° Peut être mise sous la forme*

$$P_{jk}(s, t) = \sum_l a_{jl}(s) \frac{A_{kl}(t)}{d(t)},$$

où $A_{kl}(t)$ est le coefficient de $a_{kl}(t)$ dans le développement du déterminant $d(t)$ des $a_{jl}(t)$ et où les $a_{jl}(s)$ sont des fonctions continues et dérivables choisies arbitrairement parmi celles qui satisfont aux conditions suivantes : leur déterminant $d(s)$ reste positif, les $a_{j1}(s)$ restent égaux à 1, et l'on a

$$(5) \qquad \sum_l a'_{jl}(t) A_{kl}(t) \begin{cases} \leqq 0 & \text{pour} \quad j \neq k, \\ \geqq 0 & \text{pour} \quad j = k. \end{cases}$$

Par exemple, pour $r = 2$, on pose

$$a_{12}(s) = -a(s), \qquad a_{22} = b(s) \qquad [\text{avec } a_{11}(s) = a_{21}(s) = 1];$$

on voit qu'on doit avoir

$$d(s) = a(s) + b(s) > 0$$

et les conditions (27) deviennent ici

$$a'(s) \geqq 0, \qquad b'(s) \geqq 0,$$

c'est-à-dire que les fonctions $a(s)$, $b(s)$ sont non décroissantes et de somme positive. Ce sont les conditions déjà trouvées (p. 242).

Cas homogène.

Une solution très générale. — On se trouve dans le cas dit, souvent, cas homogène quand on ne considère parmi les solutions des équations fonctionnelles (I′), (L′) que celles qui ne dépendent de s et de t que par l'intermédiaire de $t-s$, $\varphi_{jk}(s,t) = \Phi_{jk}(t-s)$.

On peut induire facilement l'existence de solutions continues très générales de cette forme, en observant qu'en posant $v = u - s$, $w = t - u$, l'équation (I′) prend la forme suivante

$$\Phi_{jk}(v+w) = \sum_{i=1}^{r} \Phi_{ji}(v)\,\Phi_{ik}(w). \tag{I''}$$

Si donc on pose, pour n entier, $a_{jk}^{(n)} = \Phi_{jk}(n)$, on a l'équation d'itération

$$a_{jk}^{(n+m)} = \sum_{i} a_{ji}^{(n)}\, a_{ik}^{(m)}, \tag{28}$$

dont la solution générale, quand le déterminant des $\Phi_{jk}(1)$ est $\neq 0$, est (Note A, p. 256) de la forme

$$a_{jk}^{(n)} = \sum_{g} e^{n\lambda_g}\, \mathrm{R}_{jkg}(n), \tag{29}$$

où les $\mathrm{R}_{jkg}(n)$ sont des polynomes en n de degrés $< r$ et où les $\sigma_g = e^{\lambda_g}$ sont des constantes distinctes, indépendantes de j, k et en nombre $\leq r$.

Les $\mathrm{R}_{jkg}(n)$ ne peuvent d'ailleurs être arbitrairement choisies. Il faut que la condition (28) soit vérifiée. D'où

$$\sum_{g} e^{(n+m)\lambda_g}\, \mathrm{R}_{jkg}(n+m) = \sum_{g} e^{n\lambda_g} \sum_{i} \mathrm{R}_{jig}(n) \sum_{g'} e^{m\lambda_{g'}}\, \mathrm{R}_{ikg'}(m).$$

Les σ_g étant distincts et les R des polynomes, on montre dans tous les Traités d'Analyse que cette égalité ne peut avoir lieu pour tout

entier n assez grand que si

$$e^{m\lambda_g} R_{jkg}(n+m) = \sum_{g'} e^{m\lambda_{g'}} \sum_i R_{jig}(n) R_{ikg'}(m),$$

d'où, ceci ayant lieu pour tout m assez grand, la condition pour les R.

$$\sum_i R_{jig}(n) R_{ikg'}(m) = \begin{cases} R_{jkg}(n+m) & \text{si } g = g', \\ 0 & \text{si } g \neq g', \end{cases}$$

condition suffisante en même temps que nécessaire pour assurer que l'expression (29) vérifie la condition (28).

Mais l'expression (29) garde un sens quand n est remplacé par un nombre v entier ou non. Dès lors, si les σ_g sont des constantes distinctes indépendantes de j et k et en nombre $\leq r$ et si les $R_{jkg}(v)$ sont des polynomes en v de degrés $\leq r$, vérifiant les conditions

$$(30) \qquad \sum_i R_{jig}(v) R_{ikg'}(w) = \begin{cases} R_{jkg}(v+w) & \text{pour } g' = g, \\ 0 & \text{pour } g' \neq g, \end{cases}$$

l'expression

$$(31) \qquad \Phi_{jk}(v) = \sum_g e^{v\sigma_g} R_{jkg}(v)$$

fournit un système de solutions continues très général de l'équation fonctionnelle (I''). Pour que ce système vérifie en outre la condition (L') qui est ici $\Phi_{jk}(0) = \delta_{jk}$, il faut et il suffit qu'on ait

$$(32) \qquad \sum_g R_{jkg}(0) = \delta_{jk}$$

On a vu, page 232, que si (L') a lieu, $D(s, t)$ reste $\neq 0$, et, par suite, l'hypothèse faite un peu plus haut que le déterminant des $\Phi_{jk}(1)$ est $\neq 0$, se trouve, dans ce cas, réalisée d'elle-même.

La solution à dérivées continues la plus générale. — Quand on ne cherche que celles des solutions de (I'') qui sont dérivables et à dérivées continues, on peut montrer que la solution précédente est la plus générale en ayant recours aux équations différentielles (9), page 205, de Kolmogoroff. Comme celui-ci l'a signalé, la formule (7), page 205, qui définit $U_{ik}(t)$, montre que si $P_{ik}(s, t)$ ne dépend que de $t-s$, $U_{ik}(t)$

est constant. Il en est de même pour les $x_{k}(t)$ définis à partir de solutions $\varphi_{jk}(s, t) = \Phi_{jk}(t - s)$ des équations (I''), (L') ne satisfaisant pas nécessairement à (T'), (P'). Dès lors les solutions cherchées sont solutions d'équations différentielles canoniques linéaires et homogènes à coefficient constants, qu'on peut écrire, en posant $v = t - s$

$$(33) \qquad \frac{dx_k}{dv} = \sum_{j=1}^{r} a_{kj} x_j(v) \qquad (k = 1, \ldots, r).$$

On trouvera dans tout Traité d'Analyse l'expression générale de ces solutions ; en prenant en particulier celles qui vérifient (I'') et (L'), on voit ainsi qu'elles sont de la forme (31) où les $R_{jkg}(v)$ sont des polynomes vérifiant les conditions (30) et (32).

Cas des probabilités. — L'étude du comportement de $\Phi_{jk}(n)$ quand n est entier (*voir* Note V) montre que : 1° si les $\Phi_{jk}(n)$ sont bornés quand n croît, les σ_g sont tous en module ≤ 1 (c'est-à-dire la partie réelle de chaque λ_g est ≤ 0), et tout $R_{jkg}(n)$ correspondant à un σ_g de module égal à 1 se réduit à une constante ; 2° si la condition (T) est vérifiée, l'un des σ_g est égal à 1 ; enfin, 3° si, en outre, la condition (P) est vérifiée, les σ_g de module égal à 1 sont racines d'une même équation binome $\sigma^N = 1$.

Dès lors, quand les conditions (I), (L), (T), (P) sont vérifiées quel que soit v, entier ou non, ces propriétés des σ_g et des R_{jkg} subsistent *a fortiori* et l'on en conclut que l'on peut écrire $P_{jk}(t - s)$ sous la forme

$$(34) \qquad P_{jk}(v) = H_{jk} + \sum_{g} e^{\frac{2\pi i m_g}{N} v} R_{jkg} + \varepsilon_{jk}(v),$$

où H_{jk}, R_{jkg}, N et m_g sont des constantes, N et m_g des entiers et où $\varepsilon_{jk}(v)$ converge exponentiellement vers zéro quand $v \to +\infty$.

Mais ici se place une simplification supplémentaire signalée d'abord par MM. Kryloff et Bogoliouboff dans le cas d'une suite continue d'états et dont M. Dœblin m'a fait observer l'application au cas actuel : c'est que la partie périodique $\sum_g$ de l'expression (34) des $P_{jk}(v)$ doit être supprimée. On le voit intuitivement en observant que pour tout μ positif et tout entier n, si l'on pose $a_{jk}^{(n)} = P_{jk}(\mu n)$, cette fonc-

tion de n satisfait à l'équation d'itération (28) et aux conditions (T) et (P). Donc $P_{jk}(\mu n)$ est une fonction de l'entier n qui a une période asymptotique (entière) N_μ. Or, cela n'est possible quel que soit μ positif que si la partie périodique de P_{jk} se réduit à une constante, comme nous voulions le prouver

Pour bien établir cette impossibilité, on peut opérer ainsi. Pour tout w positif, entier ou non, les $P_{jk}(\mu w)$ sont solutions d'équations différentielles analogues à (33), avec w pour variable et des coefficients différents. Ce sont donc des expressions de la forme, analogue à (31)

$$P_{jk}(\mu w) = \sum_h e^{w\lambda'_h} R'_{jkh}(w). \tag{35}$$

et en prenant w entier, le raisonnement fait plus haut montre que si $e^{\lambda'_h}$ est de module 1, $\lambda'_h = 2\pi i \rho'_h$ où ρ'_h est rationnel. Or, on tire de (35)

$$P_{jk}(v) = \sum_h e^{\frac{v}{\mu}\lambda'_h} R'_{jkh}\left(\frac{v}{\mu}\right) = \sum_g e^{v\lambda_g} R_{jkg}(v)$$

On sait qu'alors tout λ_g est égal à l'un des $\frac{\lambda'_h}{\mu}$. Si l'assertion à démontrer était fausse, l'un au moins des λ_g serait de la forme $2\pi i \rho_g$, où ρ_g est rationnel et non nul.

Alors $\lambda'_h = 2\pi i \mu\rho_g$ où $\mu\rho_g$ est réel et $\neq 0$, donc $e^{\lambda'_h}$ est de module 1, et, par suite,

$$2\pi i \rho_g = \lambda_g = \frac{\lambda'_h}{\mu} = \frac{2\pi i \rho'_h}{\mu},$$

d'où $\mu = \frac{\rho'_h}{\rho_g}$ = nombre rationnel. Il suffira donc de prendre le nombre positif arbitraire μ égal à un nombre irrationnel pour arriver à la contradiction annoncée.

Ainsi, dans le cas homogène, les solutions continues et dérivables du système de conditions (I), (L), (T), (P) *rentrent dans le cas non oscillant*, c'est-à-dire sont de la forme

$$P_{jk}(v) = P_{jk} + \varepsilon_{jk}(v),$$

où P_{jk} est une constante et où $\varepsilon_{jk}(v)$ converge vers zéro. Plus précisément, $\varepsilon_{jk}(v)$ converge exponentiellement vers zéro quand $v \to +\infty$

et même

$$z_{jk}(v) = \sum_g{}' e^{\lambda_g v} \cdot R_{jkg}(v)$$

où la sommation $\sum_g'$ n'est étendue qu'aux $\lambda_g \neq 0$, avec $|e^{\lambda_g}| < 1$ et où les $R_{jkg}(v)$ sont des polynomes en v.

Un calcul direct établira plus loin aisément que, dans le cas de $r = 1$ ou 2, cette expression reste valable pour toutes les solutions continues, c'est-à-dire que celles-ci sont nécessairement dérivables. D'après M. Dœblin [5], il en est encore de même pour r entier quelconque.

Du fait que $z_{jk}(v)$ converge exponentiellement vers zéro, on déduit facilement que $P_{jk}(v)$ *converge « en moyenne intégrale »* vers P_{jk}, en ce sens qu'on a

$$\lim_{T \to +\infty} \left\{ \frac{1}{T} \int_0^T P_{jk}(v)\, dv \right\} = P_{jk}$$

Il résulte de tout ce qui précède que pour trouver la limite de $P_{jk}(v)$, pour discerner si l'on est dans le cas régulier ou simplement non oscillant, il suffit de résoudre les problèmes correspondants pour la fonction de *l'entier* n, $P_{jk}(n)$ obtenu en itérant n fois $P_{jk}(1)$. Et nous savons comment résoudre ces problèmes connaissant seulement les r^2 nombres $P_{jk}(1)$.

Cas de $r = 1$ ou 2. — Dans des cas simples comme ceux de $r = 1$ ou 2, on peut déterminer par un raisonnement direct les solutions continues de (F_r) ne dépendant que de $t - s$ sans faire appel au cas d'une suite discontinue d'épreuves ou à la réduction à des équations différentielles.

1° $r = 1$. Considérons d'abord le cas de $r = 1$. On a vu que pour le système

$$D(s, t) = D(s, u)\, D(u, t), \qquad D(s, s) = 1,$$

la solution la plus générale jamais nulle, > 0, est de la forme $D(s, t) = \frac{A(s)}{A(t)}$, où $A(t)$ est toujours positive. Si $D(s, t)$ est une

fonction $u(t-s)$ de $t-s$, on aura pour $t=s+s'$

$$\Lambda(s+s') = \frac{\Lambda(s)}{u(s')}. \tag{36}$$

D'où

$$\frac{\Lambda(s)}{u(s')} = \frac{\Lambda(s')}{u(s)} \quad \text{ou} \quad \Lambda(s)\,u(s) = \Lambda(s')\,u(s')$$

pour toute valeur positive de s'. Dès lors $\Lambda(s)\,u(s)$ a une valeur constante $\frac{1}{k} > 0$ indépendante de s, d'où

$$\Lambda(s+s') = k\,\Lambda(s)\,\Lambda(s'),$$

et, en posant $f(s) = \log[k\,\Lambda(s)]$, on aura

$$f(s+s') = f(s) + f(s'). \tag{37}$$

Quand on suppose $D(s,t)$ continue, $\Lambda(s)$ est continue et >0, donc $f(s)$ est continue. Comme on sait, toute solution $f(s)$ de (37) est alors de la forme Us où U est une constante, d'où

$$k\,\Lambda(s) = e^{Us},$$

et alors

$$D(s,t) = e^{U(s-t)}.$$

Dans le cas où $D(s,t)$ est bornée pour $t-s \geqq 0$, par exemple si $D(s,t)$ est une probabilité, U doit être $\geqq 0$ et alors, on aura $0 < D(s,t) \leqq 1$.

2° $r=2$. Les solutions continues de (I), (L), (T), (P) dans le cas de $r=2$, sont de la forme (22), page 242.

Si les $\varphi_{ik}(s,t)$ sont des fonctions de $s'=t-s$, on a

$$\frac{d(s)}{d(t)} = \varphi_{11}(s,t) + \varphi_{22}(s,t) - 1 = u(t-s),$$

d'où

$$d(s+s') = \frac{d(s)}{u(s')},$$

identité analogue à (36). On en tire de même

$$k\,d(s) = e^{\lambda s}, \tag{38}$$

où λ est une constante et où k est une constante >0. On a alors

$$\frac{a(t)-a(s)}{d(t)} = \varphi_{12}(s,t) = \rho(t-s),$$

d'où

$$a(t) = a(s) + e^{\lambda t}\,\frac{v(t-s)}{\lambda},$$

ou

$$a(s + s') = a(s) + e^{\lambda s}\,\varpi(s').$$

On pourrait raisonner ensuite très simplement en supposant a et ϖ dérivables. Mais ce n'est pas indispensable. On voit qu'on a $\varpi(0) = 0$ et

$$(39)\qquad a(s) + e^{\lambda s}\,\varpi(s') = a(s') + e^{\lambda s'}\,\varpi(s),$$

et, par suite, pour $s' = 0$

$$(40)\qquad \varpi(s) = [a(s) - a_0] \qquad \text{avec} \quad a_0 = a(0).$$

D'où, en portant ϖ d'après (40) dans (39),

$$a(s) + e^{\lambda s}[a(s') - a_0] = a(s') + e^{\lambda s'}[a(s) - a_0]$$

$$[a(s) - a_0][e^{\lambda s'} - 1] = [a(s') - a_0][e^{\lambda s} - 1],$$

ou, si $\lambda \neq 0$,

$$\frac{a(s) - a_0}{e^{\lambda s} - 1} = \frac{a(s') - a_0}{e^{\lambda s'} - 1}.$$

La valeur commune de ces rapports est une quantité c indépendante de s et de s'. On a donc

$$(41)\qquad a(s) - a_0 = c(e^{\lambda s} - 1),$$

et d'après (38)

$$(42)\qquad b(s) = d(s) - a(s) = \frac{1}{\lambda}e^{\lambda s} - ce^{\lambda s} - a_0 + c = c'(e^{\lambda s} - 1) + b_0$$

avec $c + c' = \frac{1}{\lambda} > 0$.

D'où, d'après (22), (41) et (42),

$$P_{11}(s, t) = 1 - \frac{c}{c + c'}\,\frac{e^{\lambda t} - e^{\lambda s}}{e^{\lambda t}}, \qquad P_{21}(s, t) = \frac{c'}{c + c'}\,\frac{e^{\lambda t} - e^{\lambda s}}{e^{\lambda t}};$$

$$P_{12}(s, t) = \frac{c}{c + c'}\,\frac{e^{\lambda t} - e^{\lambda s}}{e^{\lambda t}}, \qquad P_{22}(s, t) = 1 - \frac{c'}{c + c'}\,\frac{e^{\lambda t} - e^{\lambda s}}{e^{\lambda t}}.$$

ou, puisque $c + c' > 0$, en posant

$$p = \frac{c}{c + c'}, \qquad q = \frac{c}{c + c'};$$

$$(43) \qquad \begin{cases} P_{11}(s, t) = p + q e^{\lambda(s-t)}, & P_{21}(s, t) = p[1 - e^{\lambda(s-t)}], \\ P_{12}(s, t) = q[1 - e^{\lambda(s-t)}], & P_{22}(s, t) = q + p e^{\lambda(s-t)}. \end{cases}$$

On voit que les $P_{ik}(s, t)$ sont bien des fonctions de $s - t$, qu'on a

$$\sum_k P_{ik}(s, t) = p + q = 1 \qquad \text{et} \qquad P_{ik}(s, s) = \delta_{ik}$$

Reste la condition (P). On a

$$P_{12}(s, t) + P_{21}(s, t) = 1 - e^{\lambda(s-t)}$$

qui doit être $\geqq 0$ pour $t > s$, il faut donc que λ soit $\geqq 0$, et alors pour que P_{21} et P_{12} restent $\geqq 0$, il faut que p et q soient $\geqq 0$. Réciproquement si λ, p, q sont $\geqq 0$ et $p + q = 1$, les P_{jk} de (43) sont $\geqq 0$.

En résumé *la solution continue la plus générale*, ne dépendant que de $s - t$, des conditions (I), (T), (P), (L), est de la forme (43) où

$$\lambda \geqq 0, \qquad p \geqq 0, \qquad q \geqq 0, \qquad p + q = 1$$

Le raisonnement conduisant à (43) avait supposé $\lambda \neq 0$, pour obtenir (41). Mais si $\lambda = 0$, $d(s)$ est constant, donc aussi $a(s)$ et $b(s)$ et par suite les formules (43) restent valables.

Le cas où $\lambda = 0$ est un cas singulier non oscilllant où $P_{jk}(s, t) = \delta_{jk}$. En dehors de ce cas exceptionnel, on observe qu'*on est alors nécessairement dans le cas régulier*. Il suffirait pour cela que $P_{ik}(s, t)$ convergeât quand t croît vers une limite $P_k(s)$ indépendante du premier indice. Ici la limite est même indépendante de l'époque initiale s

$$\lim_{t \to \infty} P_{i1}(s, t) = p, \qquad \lim_{t \to \infty} P_{i2}(s, t) = q.$$

Nota — Nous devons nous contenter de signaler deux études sur des problemes nouveaux concernant le cas homogène : l'une de M. Elfving [1], parue à la fin de la correction des épreuves de ce livre, l'autre de M. Dœblin [5], en cours d'impression.

COMPLÉMENTS DE MATHÉMATIQUES PURES.

QUATRE NOTES SUR LES SYSTÈMES D'ÉQUATIONS LINÉAIRES AUX DIFFÉRENCES FINIES DU PREMIER ORDRE À COEFFICIENTS CONSTANTS

NOTE A

Comportement asymptotique des solutions dans le cas des systèmes homogènes [1]

Expression des solutions. — Pour mieux marquer la généralité de nos hypothèses, écrivons l'équation d'itération (I) de la page 24 en changeant de notation et remplaçant $p_{ik}^{(n)}$ par $a_{ik}^{(n)}$ qui sera une quantité sans relation nécessaire avec le Calcul des Probabilités.

Soit le système de r^2 relations

$$(I) \qquad a_{jk}^{(n+m)} = \sum_{i=1}^{i=r} a_{ji}^{(n)} a_{ik}^{(m)} \qquad (j, k = 1, \ldots, r).$$

Elles permettent de déterminer de proche en proche les $a_{ik}^{(n)}$ à partir des quantités $a_{ik} = a_{ik}^{(1)}$, supposées entièrement arbitraires, contrairement aux p_{ik}, au moyen des r^2 relations

$$(I_1) \qquad a_{jk}^{(n+1)} = \sum_{i=1}^{i=r} a_{ji} a_{ik}^{(n)} \qquad (j, k = 1, \ldots, r).$$

On peut alors considérer $a_{1k}^{(n)}, \ldots, a_{rk}^{(n)}$ comme des solutions d'un

---

[1] Pour les démonstrations, *voir* Fréchet [152].

système d'équations aux différences finies du premier ordre

$$\Delta x_j(n) = \sum_{i=1}^{i=r} \Lambda_{ji}\, x_i(n) \qquad (j = 1, \quad, r), \tag{1}$$

où $\Delta x_j(n) = x_j(n+1) - x_j(n)$, et où l'on a posé

$$\Lambda_{ji} = a_{ji} \quad \text{si } j \neq i \qquad \text{et} \qquad \Lambda_{jj} = a_{jj} - 1$$

ou, ce qui revient au même, du système d'équations récurrentes

$$x_j(n+1) = \sum_{i=1}^{i=r} a_{ji}\, x_i(n) \qquad (j = 1, \ldots, r). \tag{2}$$

On voit immédiatement que la solution générale de (2) s'écrit sous la forme

$$x_j(n) = \sum_{g=1}^{g=r} \mu_g\, X_{jg}(n) \qquad (j = 1, \quad, r), \tag{3}$$

où les $X_{jg}(n)$ sont r solutions linéairement indépendantes.

Le système (1) est tout à fait analogue au système de r équations différentielles linéaires et homogènes, du premier ordre et à coefficients constants, à r fonctions inconnues

$$\frac{d\,x_j(t)}{dt} = \sum_{i=1}^{i=r} \Lambda_{ji}\, x_i(t) \qquad (j = 1, \;.\;, r).$$

La solution de ce dernier système figure dans tous les Traités d'Analyse; un raisonnement tout semblable (Lublin [1]), corrigé par une réserve sur laquelle nous reviendrons quelques lignes plus loin, fournit la solution générale du système (1). D'après ce raisonnement, il suffirait de prendre, en particulier, dans (3), pour les $X_{jg}(n)$, r solutions particulières de la forme

$$Z_{jg}(x) = (s_g)^n\, \rho_{jg}(n) \qquad (g = 1, \ldots, r) \tag{4}$$

où $s_1, s_2, \ldots, s_r$ sont les r racines, distinctes ou non, de « *l'équation*

en s » du système (2), soit $\Delta(s) = 0$ avec

$$\Delta(s) = \begin{vmatrix} a_{11}-s & a_{21} & \cdot & a_{r1} \\ a_{12} & a_{22}-s & & a_{r2} \\ & \cdot & & \\ a_{1r} & a_{2r} & & a_{rr}-s \end{vmatrix},$$

et où $c_{jg}(n)$ est un polynome en n de degré inférieur à l'ordre de multiplicité de la racine s_g de $\Delta(s)$.

La même forme de solution avait été obtenue antérieurement par M. Konečný [1, p. 7] comme conséquence de propriétés des matrices dues à Frobenius [1 - 4] et à M. Perron. Ce dernier a même donné une expression entièrement explicite de $a_{jk}^{(n)}$, [Perron, 1, formule (15)]. Cette expression, qui exige le calcul de déterminants, est assez peu maniable, mais elle est valable pour tout entier n. L'expression obtenue par MM. Konečný et Lublin est plus commode pour notre but, bien qu'il reste à y déterminer les coefficients, mais elle comporte un cas d'exception sur lequel nous reviendrons quelques lignes plus loin.

La forme générale de la solution qui va en être déduite n'est pas, en effet, toujours valable quel que soit l'entier n, comme l'ont observé M. Čech et M. Dœblin, et comme le montreront les exemples de la page 273. On verra, page 268, que cette forme générale est au contraire valable quel que soit l'entier n, si $\Delta(0) \neq 0$, et quand $\Delta(0) = 0$, elle est valable pour toutes les valeurs de n supérieures à l'ordre de multiplicité de zéro comme racine de $\Delta(s)$.

En appelant $\sigma_1, \ldots, \sigma_q$, $(q \leq r)$, *les racines distinctes de* $\Delta(s)$, on voit que, pour $n > r$, *les* $a_{jk}(n)$ *sont de la forme*

$$a_{jk}^{(n)} = \sum_{g=1}^{g=q} (\sigma_g)^n w_{jkg}(n), \tag{5}$$

où $w_{jkg}(n)$ *est aussi un polynome en* n *de degré inférieur à l'ordre de multiplicité de* σ_g.

Nous avons en vue l'application ultérieure de nos résultats concernant le système (I_1) au cas où les $a_{jk}^{(n)}$ sont des probabilités, nécessairement comprises entre 0 et 1. Il est donc naturel de chercher à distinguer pour le cas général, sous quelles conditions les $a_{jk}^{(n)}$ restent

bornés quel que soit n. Nous dirons qu'on est alors dans « *le cas borné* ».

La condition nécessaire et suffisante pour qu'on soit dans le cas borné est : 1° que l'équation « en s », $\Delta(s)=0$, n'ait que des racines ≤ 1, en module; 2° que pour toute racine éventuelle s_g de module 1, les polynomes correspondants $\varpi_{jkg}(n)$ se réduisent tous à des constantes par rapport à n.

Dans ce même cas borné, les σ_g vont donc se partager en deux groupes suivant que $|\sigma_g|=1$ ou $|\sigma_g|<1$. Appelons $\varpi_{jk}(n)$, $\varepsilon_{jk}(n)$ la somme des termes correspondant respectivement à chacun de ces deux groupes dans l'expression (5) de $a_{jk}^{(n)}$. Il peut d'ailleurs arriver qu'une de ces deux sommes se réduise à zero. Il est clair que les $\varepsilon_{jk}(n)$ convergent vers zéro; de sorte que $a_{jk}^{(n)}$ se comporte de la même manière quand n croît, que $\varpi_{jk}^{(n)}$, en ce sens que leur différence tend vers zéro. Or, $\varpi_{jk}^{(n)}$ a, d'après ce qui précède, une expression très simple, étant la somme de termes de la forme $\varpi_{jkg}e^{in\varphi_g}$ où ϖ_{jkg} et le nombre réel φ_g sont indépendants de n. A ce titre, cette somme $\varpi_{jk}^{(n)}$, qui peut être appelée une « fonction trigonométrique de n », est une fonction « quasi-périodique » de n au sens de Bohl, d'Esclangon (cas très particulier des fonctions « presque périodiques » de H. Bohr qui sont des *séries* illimitées de tels termes)

Convergence en moyenne. — Une des propriétés les plus importantes des fonctions quasi-périodiques s'exprime plus simplement quand on utilise une notion due à Cesaro :

Nous dirons qu'une suite de nombres $u_1, u_2, \ldots, u_n, \ldots$, *converge en moyenne arithmétique* (*ou au sens de Cesaro*) vers λ quand la moyenne arithmétique $\frac{u_1+\ldots+u_n}{n}$ tend vers λ au sens ordinaire Pour les raisons exposées page 71, on peut alors dire que λ est la limite généralisée de u_n Observons d'ailleurs — ce qui sera utile plus loin — que λ est alors aussi la limite généralisée de $u_\nu, u_{\nu+1}, \ldots, u_{\nu+n}, \ldots$, quel que soit ν fixe, et réciproquement Ceci étant : un calcul immédiat montre que toute fonction trigonométrique de n (de la forme précisée ci-dessus) converge en moyenne quand n croît. Dès lors, il en est de même des $\varpi_{jk}^{(n)}$ et, par suite, aussi des $a_{jk}^{(n)}$.

Finalement : *dans le cas borné, les $a_{jk}^{(n)}$ convergent toujours, au moins au sens de Cesaro, vers des limites généralisées* Π_{jk}.

En posant

$$(6)\qquad \mathrm{H}_{jk}^{(n)} = \frac{1}{n}(a_{jk}^{(1)} + \ldots + a_{jk}^{(n)}),$$

les différences $\mathrm{H}_{jk}^{(n)} - \mathrm{H}_{jk}$ sont donc infiniment petites avec $\frac{1}{n}$. On peut même en évaluer l'ordre : cet ordre est au moins égal à celui de $\frac{1}{n}$. On peut énoncer ce résultat sous une forme équivalente qui est parfois plus commode : *dans le cas borné, les sommes partielles des séries*

$$\sum_{n=1}^{\infty} |a_{jk}^{(n)} - \mathrm{H}_{jk}|$$

sont bornées dans leur ensemble.

On peut exprimer plusieurs des résultats précédemment énoncés en disant que les $a_{jk}^{(n)}$ sont des fonctions « asymptotiquement » quasi périodiques de n. Ce résultat incite à examiner s'il serait possible de supprimer dans cette phrase le mot quasi. On voit facilement que *dans le cas borné, la condition nécessaire et suffisante pour que les $a_{jk}^{(n)}$ soient toutes des fonctions asymptotiquement périodiques de n* — c'est-à-dire soient chacune somme d'une fonction périodique de n et d'une quantité tendant vers zéro avec $\frac{1}{n}$ — *est que les racines de module* 1 *de* $\Delta(s)$, s'il en existe, *soient racines d'une même équation binome*. Et alors, *le degré de cette équation binome est une « période asymptotique » commune aux $a_{jk}^{(n)}$.*

En particulier, *pour que dans le cas borné, les $a_{jk}^{(n)}$ soient toutes des fonctions périodiques de n, il faut et il suffit que toutes les racines de $\Delta(s)$ soient aussi racines d'une même équation binome.* Et alors *le degré de cette équation binome est une période asymptotique commune aux $a_{jk}^{(n)}$.*

Valeurs des limites généralisées des $a_{jk}^{(n)}$. — En vertu des relations (I_1), *les limites généralisées* H_{jk} *des $a_{jk}^{(n)}$ vérifient les équations*

$$(7)\qquad \mathrm{H}_{jk} = \sum_l \mathrm{H}_{jl} a_{lk} = \sum_l a_{jl} \mathrm{H}_{lk} = \sum_l \mathrm{H}_{jl} \mathrm{H}_{lk}.$$

De sorte que $\Pi_{1k}, \ldots, \Pi_{rk}$ sont solutions du système

$$(8) \qquad y_j = \sum_i a_{ij} y_i \qquad (j = 1, \ldots, r)$$

On en conclut : quand, dans le cas borné, l'unité n'est pas racine de $\Delta(s)$, les $a_{jk}^{(n)}$ convergent, au moins au sens de Cesaro, tous vers zéro.

Quand $\Delta(1) = 0$, alors les équations (8) ont au moins un système de solutions non toutes nulles, et ils auront en général un certain nombre $\rho \leqq r$ de systèmes linéairement indépendants $S_g (g = 1, \ldots, \rho)$ de solutions

$$(S_g) \qquad X_{1g}, \quad \ldots, \quad X_{rg}$$

Dans le cas borné, on peut donner *l'expression suivante des limites généralisées* Π_{jk}.

$$\Pi_{jk} = \sum_{g=1}^{g=\rho} X_{jg} Y_{kg},$$

où les $Y_{1g}, \ldots, Y_{rg}$ *sont un système* S'_g *de solutions des équations*

$$(9) \qquad y_k = \sum_i y_i a_{ik} \qquad (k = 1, \ldots, r)$$

associées à (8), *les systèmes* $S'_1, \ldots S'_r$ *étant bien déterminés — quand on a choisi arbitrairement les systèmes linéairement indépendants* $S_1, \ldots, S_r$ — *par la condition que les systèmes* S_g, S'_g *soient biorthonormés, c'est-à-dire tels que*

$$(10) \qquad \sum_i X_{ig'} Y_{ig} = \delta_{gg'} \qquad \text{avec} \quad \delta_{gg'} = \begin{cases} 1 & \text{si } g = g', \\ 0 & \text{si } g \neq g'. \end{cases}$$

L'étude du cas particulier des probabilités invite à poser quelques questions dont nous allons donner les réponses :

Dans le cas borné, la condition nécessaire et suffisante pour que les limites généralisées Π_{jk} *des* $a_{jk}^{(n)}$ *soient indépendantes du premier indice, j, sans être toutes nulles est :* 1° *que la condition*

$$(T_1) \qquad \sum_i a_{ji} = 1$$

soit vérifiée, 2° *que l'unité soit racine simple de l'équation en s.* On observera que $\Delta(1)$ étant nul quand (T_1) est vérifiée, la condition 2° ne porte que sur l'ordre de multiplicité de la racine 1.

On en déduit que :

Dans le cas borné, pour que les limites généralisées H_{jk} *des* a''_{jk} *soient indépendantes de* j *et de* k (c'est-à-dire toutes égales) *sans être toutes nulles, il faut et il suffit :* 1° *que les deux conditions* (T_1) *et*

$$(T_1) \qquad \sum_i a_{ji} = 1$$

soient vérifiées, 2° *que l'unité soit racine simple de l'équation en* s.

Alors *la valeur commune des* H_{jk} *est égale à* $\frac{1}{r}$.

La condition 2° peut être avantageusement remplacée par une autre conduisant au calcul des H_{jk} quand ceux-ci sont seulement indépendants de i. On obtient ainsi le résultat suivant

Dans le cas borné, pour que les limites généralisées H_{ji} *des* a''_{jk} *soient indépendantes du premier indice sans être toutes nulles, il faut et il suffit :* 1° *que la condition*

$$(T_1) \qquad \sum_i a_{ji} = 1$$

soit réalisée; 2° *que le système d'équations*

$$(E) \qquad \begin{cases} y_k = \sum_i a_{ik} y_i & (k = 1, \ldots, r), \\ \sum_i y_i = 1 \end{cases}$$

ait un système unique de solutions. Et, dans ce cas, *les limites* $H_i (= H_{1i} = \ldots = H_{ri})$ *ont précisément pour valeurs celles de la solution unique de* (E).

Observons, d'ailleurs, que les $r+1$ équations (E) à r inconnues ne sont pas indépendantes quand (T_1) a lieu, la somme des r premières se réduisant alors à une identité.

Passons enfin à la convergence ordinaire :

Dans le cas borné, pour que les $a_{jk}^{(n)}$ convergent tous au sens ordinaire, il faut et il suffit que $\Delta(s)$ n'ait aucune racine qui soit à la fois $\neq 1$ et de module 1. Quand cette condition est remplie, non seulement *le produit* $n[\Pi_{jk}^{(n)} - \Pi_{jk}]$ est borné, comme on l'a vu plus haut, mais, en outre, *il a une limite* s_{jk} (1). Cette limite est la somme de la série absolument convergente

$$s_{jk} = \sum_{n=1}^{n=+\infty} [a_{jk}^{(n)} - \Pi_{jk}]$$

Pour calculer s_{jk} au moyen de cette formule, il faudrait calculer tous les $a_{jk}^{(n)}$. Pour éviter ce calcul, on prouve d'abord que $s_{h1}, \ldots, s_{hr}$ *sont solutions du système*

$$(E_h) \quad \begin{cases} \sum_j z_j a_{jk} - z_k = \Pi_{hk} - a_{hk} & (k = 1, \ldots, r), \\ \sum_l z_l \Pi_{lk} = 0 \end{cases}$$

Pour h déterminé, on a $r+1$ équations à r inconnues; mais les r premières ne sont pas indépendantes, car en les multipliant par $\Pi_{1l}, \ldots, \Pi_{rl}$, et en ajoutant, on obtient une identité

Dans le cas particulier, où les limites généralisées sont indépendantes du premier indice et non toutes nulles, le système (E_h) se simplifie et prend la forme

$$(E'_h) \quad \begin{cases} \sum_j z_j a_{jk} - z_k = \Pi_k - a_{hk} & (k = 1, \ldots, r), \\ \sum_l z_l = 0 \end{cases}$$

Si, par exemple, on ne tient pas compte dans ce système, de même que dans le système (E) de la page précédente, de la premiere équation, on a deux systèmes linéaires qui ont le même déterminant des coefficients, et dont le second n'a, comme on l'a vu, qu'un seul système de solutions. Donc :

(1) *Voir* la Note B pour l'extension de la définition de s_{jk} au cas borné le plus général.

Dans le cas régulier, — c'est-à-dire quand les $a_{hk}^{(n)}$ convergent au sens ordinaire vers des limites H_k *indépendantes du premier indice, h —, on peut déterminer les* $s_{h1}, \ldots, s_{hl}$ *au moyen du système* (E'_h), *d'une manière unique* et sans avoir à effectuer d'itération.

Cas symetrique. — Quand a_{jk} est une fonction réelle symétrique de j et de k, on sait que les racines de $\Delta(s)$ sont toutes réelles. De plus, il est clair que, dans ce cas, les conditions (T_1) et (T'_1) ne peuvent être réalisées qu'en même temps.

Une racine de $\Delta(s)$ de module 1 ne peut être réelle que si elle est égale à $+1$ ou à -1. Donc :

Dans le cas borné symétrique, ou bien -1 *n'est pas racine de* $\Delta(s)$, *et alors les* $a_{jk}^{(n)}$ *convergent au sens ordinaire, ou bien* -1 *est racine de* $\Delta(s)$, *et* $a_{jk}^{(n)}$ *est une fonction asymptotiquement périodique de n, de période* 2. Enfin, si les limites généralisées H_{jk} sont indépendantes de l'un des indices j, k, elles sont indépendantes de l'autre et leur valeur commune est zéro ou $\frac{1}{l}$.

Cas des racines simples. — On peut déterminer une expression complète des solutions $a_{jk}^{(n)}$ quand les racines de $\Delta(s)$ sont toutes simples et $\neq 0$.

Dans ce cas, les formules (3) et (4) de la page 257 donnent

$$x_j(n) = \sum_{g=1}^{g=l} \mu_g v_{jg}(s_g)^n,$$

et, par suite, les $a_{jk}^{(n)}$ sont de la forme

$$a_{jk}^{(n)} = \sum_g v_{jg}\,\mu_{kg}(s_g)^n. \tag{11}$$

Pour déterminer les v et les μ, quand les a_{jk} sont donnés, substituons dans (I_1) ; on aura

$$\sum_g \mu_{kg}\left[s_g v_{jg} - \sum_g a_{jl} v_{lg}\right](s_g)^n = 0. \tag{12}$$

Pour que (I_1) soit vérifié, il suffit donc que les v vérifient les

équations

$$(13) \qquad s_g v_{jg} = \sum_{i=1}^{i=r} a_{ji} v_{ig} \qquad (j=1, \ldots, r)$$

Et puisque (I_1) détermine de façon unique les $a_{jk}^{(n)}$ connaissant les a_{jk}, on obtiendra, connaissant les a_{jk}, les $a_{jk}^{(n)}$ sous la forme (11), en assujettissant les v à vérifier les conditions (13), et les v étant choisis, en prenant des μ vérifiant la condition déduite de (11) pour $n=1$:

$$(14) \qquad a_{jk} = \sum_{g=1}^{g=r} v_{jg} \mu_{kg}$$

La détermination des v, μ est donc réduite à la résolution *successive* de deux systèmes (13), (14) du premier degré, l'un en v, l'autre en μ. *Tout système de solutions en* v *et* μ *des systèmes* (13), (14) *fournira l'expression* (11) *des* $a_{jk}^{(n)}$.

On peut ajouter des précisions utiles à ce résultat. On sait qu'une expression de la forme $\sum_{g=1}^{g=r} M_g(s_g)^n$ ne peut être nulle quel que soit l'entier n, quand les s_g sont tous distincts, que si tous les M_g sont nuls.

Pour vérifier (I_1), non seulement il est suffisant, mais on voit donc qu'il est, en outre, nécessaire qu'on ait

$$\mu_{kg}\left[s_g v_{jg} - \sum_i a_{ji} v_{ig}\right] = 0$$

Or, d'après (14), le déterminant D des a_{jk} est égal au produit de celui des v_{jg} par celui des μ_{kg}. Donc : si (13) est vérifié, ou bien il existe une valeur de g telle que $\mu_{1g} = \ldots = \mu_{rg} = 0$, et alors le déterminant des μ étant nul, D serait nul et, puisque $D = \Delta(0)$, $\Delta(s)$ aurait une racine nulle ; ou bien si $\Delta(s)$ n'a pas de racine nulle, les équations (13) seront nécessairement vérifiées, pour chaque valeur de g, par au moins un système de solutions $v_{1g}, \ldots, v_{rg}$. Celles-ci ne seront pas toutes nulles puisque le déterminant des v n'est pas nul si $D \neq 0$.

En répétant le même raisonnement sur l'équation

$$(\mathrm{I}_2) \qquad a_{jk}^{(n+1)} = \sum_i a_{ji}^{(n)} a_{ik},$$

on verrait que, si $D \neq 0$, les quantités $\mu_{1g}, \ldots, \mu_{rg}$ forment nécessairement pour chaque valeur de g un système de solutions non toutes nulles des équations

$$(15) \qquad s_g \mu_{kg} = \sum_i \mu_{ig} a_{ik} \qquad (k = 1, \ldots, r)$$

D'ailleurs, que D soit nul ou non, il est visible que les systèmes (13), (15) ont chacun, pour chaque valeur de g, un système de solutions non toutes nulles, puisque le déterminant des coefficients des inconnues est $\Delta(s_g) = 0$. On obtiendra un autre système de solutions non toutes nulles en multipliant les valeurs du premier par un même facteur arbitraire $\neq 0$ et pouvant varier avec g. On obtient, d'ailleurs, ainsi tous les systèmes de solutions non toutes nulles. Sans quoi, tous les mineurs d'ordre $r-1$ de $\Delta(s_g)$ seraient nuls, et $\Delta'(s_g)$ qui est la somme de tels mineurs serait nul alors que s_g est supposé racine simple.

Comme on sait que $a_{jk}^{(n)}$ *peut* être mis sous la forme (11), on en conclut finalement que si $D \neq 0$, il existe certainement au moins un système de solutions (non toutes nulles pour g donné), v_{jg}, μ_{kg} des systèmes (13), (14) et (15).

Considérons alors un système arbitraire de solutions non toutes nulles V_{jg} de (13). On aura $v_{jg} = l_g V_{jg}$, avec $l_g \neq 0$, d'où

$$a_{jk}^{(n)} = \sum_g V_{jg} l_g \mu_{kg} (s_g)^{n-1}, \qquad a_{jk} = \sum_g V_{jg} l_g \mu_{kg},$$

$$s_g l_g \mu_{kg} = \sum_i l_g \mu_{ig} a_{ik}$$

Dès lors, on voit qu'il existe un système de solutions $\Lambda_{kg} = l_g \mu_{kg}$, non toutes nulles pour g donné, des équations (14), (15) telles que $a_{jk}^{(n)}$ puisse être mis sous la forme (11). Ce système est unique, car le déterminant des coefficients des inconnues Λ_{kg}, pour k donné, dans (14), est le déterminant des V_{jg} qui est égal à $\frac{1}{l_1 \ldots l_r}$ multiplié par le déterminant des v_{jg}, lequel est certainement $\neq 0$.

En résumé, quand $D \neq 0$, pour mettre $a_{jk}^{(n)}$ sous la forme voulue, on peut proceder ainsi : on détermine pour chaque g un système quelconque de solutions non toutes nulles V_{jg} du système du premier degré (13), et cette détermination est certainement possible. Ceci fait, on prend pour les μ_{kg} le système de solutions Λ_{kg} — qui existe certainement et est bien déterminé — du système du premier degré

$$a_{jk} = \sum_{g} V_{jg} \mu_{kg} \tag{16}$$

D'ailleurs, au lieu de résoudre directement ce système, on peut procéder ainsi : on détermine pour chaque g un système de solutions non nulles Λ'_{kg} du système du premier degré homogene (15), il existe certainement un tel système de solutions On prendra alors

$$\Lambda_{kg} = l'_{g} \Lambda'_{kg},$$

et il restera à résoudre le systeme

$$a_{jk} = \sum_{g} l'_{g} V_{jg} \Lambda'_{kg},$$

par rapport aux r inconnues $l'_1, \ldots, l'_r$, système qui aura certainement un systeme unique de solutions $\neq 0$. Un changement convenable de notation conduit à des formules intéressantes dues à Frobenius [1], M. Romanovski [1]. Restant encore dans le cas où $\Delta(s)$ n'a pas de racine nulle, on peut, en changeant de notations dans (11), écrire

$$a_{jk}^{(n)} = \sum_{g} \frac{\varphi_{jg} \psi_{kg}}{(\lambda_g)^n},$$

où $\lambda_g = \frac{1}{s_g}$. Le raisonnement précédent appliqué à cette forme montre que les $\varphi_{1g}, \ldots, \varphi_{rg}$ sont un système de solutions non toutes nulles des équations

$$\varphi_{jg} = \lambda_g \sum_{i} a_{ji} \varphi_{ig} \qquad (j = 1, \ldots, r), \tag{17}$$

les $\psi_{1g}, \ldots, \psi_{rg}$, un système de solutions non toutes nulles des équations

$$\psi_{kg} = \lambda_g \sum_{i} \psi_{ig} a_{ik}, \tag{18}$$

et que les coefficients de proportionnalité de ces solutions doivent être choisis, de sorte que

$$(19) \qquad a_{ik} = \sum_g \frac{\varphi_{ig}\psi_{kg}}{\lambda_g}.$$

Enfin, les déterminants des φ_{ig} et des ψ_{kg} sont $\neq 0$.

Or, en multipliant l'équation (19) par $\varphi_{kg'}$, et ajoutant pour $k = 1, 2, \ldots, r$, on a, d'après (17),

$$\frac{1}{\lambda_{g'}}\varphi_{ig'} = \sum_g \varphi_{ig}\frac{u_{gg'}}{\lambda_g}, \qquad \text{où} \quad u_{gg'} = \sum_k \varphi_{kg'}\psi_{kg},$$

relation de la forme

$$\sum_g \varphi_{ig} w_{gg'} = 0.$$

Pour g' fixe et i variable, on a des équations linéaires et homogènes en $w_{1g'}, w_{2g'}, \ldots$, dont le déterminant des coefficients est $\neq 0$. Donc,

$$0 = w_{gg'} = \begin{cases} \dfrac{u_{gg'}}{\lambda_g} & \text{si } g \neq g', \\ \dfrac{u_{gg}-1}{\lambda_g} & \text{si } g = g'. \end{cases}$$

On a donc

$$(20) \qquad \sum_k \varphi_{kg'}\psi_{kg} = \delta_{gg'} \qquad \text{avec} \quad \delta_{gg'} = \begin{cases} 0 & \text{si } g \neq g', \\ 1 & \text{si } g = g'. \end{cases}$$

C'est-à-dire que le système des $\varphi_{kg'}$ et des ψ_{kg} est *biorthonormé* relativement au second indice [Hostinský, 14 *bis*].

En multipliant l'équation (19) par ψ_{kg}, et ajoutant pour $g = 1, \ldots, r$, un raisonnement analogue montrerait que l'on a

$$(21) \qquad \sum_g \varphi_{ig}\psi_{kg} = \delta_{ik},$$

c'est-à-dire que les φ_{ig} et les ψ_{kg} forment un système *biorthonormé* relativement au premier indice.

Supplément à la Note A.

Seconde méthode de résolution du système (2) d'équations aux différences. — Il sera parfois commode, au point de vue pratique, et

il va nous être utile au point de vue théorique, de résoudre le système (2) de la page 257, en prenant comme inconnue auxiliaire une fonction jouant un rôle analogue à celui des noyaux résolvants de Fredholm ([1]) ou à celui des fonctions génératrices de Laplace ([2]).

Considérons à cet effet la série entière en λ.

$$X_j(\lambda) = x_j(1) + \sum_{n=1}^{\infty} x_j(n+1)\lambda^n, \tag{22}$$

où $x_j(n)$ est solution de (2) Dans le cas borné, cette série entière converge pour $|\lambda| < 1$ et par suite $X_j(\lambda)$ est une fonction de λ, holomorphe à l'origine. Qu'on soit dans le cas borné ou non, si cette série converge pour $|\lambda|$ assez petit, sa somme est une fonction holomorphe à l'origine. Cette fonction intermédiaire sera utilisée pour définir les $x_j(n)$.

Elle vérifie évidemment le système d'équations linéaires

$$X_j(\lambda) = x_j(1) + \lambda \sum_{i=1}^{r} a_{ji} X_i(\lambda), \qquad (j = 1, \ldots, r)$$

qu'on peut écrire

$$\sum_{i=1}^{r} [\lambda a_{ji} - \delta_{ji}] X_i(\lambda) = -x_j \qquad (j = 1, \ldots, r) \tag{23}$$

en posant $x_j = x_j(1)$.

Dès lors la résolution du système d'équations aux différences (1) est ramenée à celle du système d'équations algébriques linéaires (23), suivie du développement des solutions en séries entières en λ.

Pour des coefficients numériques a_{ji} déterminés, il peut arriver que ces deux opérations se fassent simplement.

Dans le cas général, la solution de (23), s'obtiendra suivant la règle de Cramer. Le déterminant $D(\lambda)$ des coefficients des $X_i(\lambda)$ est égal à $(-1)^r$ pour $\lambda = 0$, c'est un polynome en λ, $D\lambda^r + \ldots + (-1)^r$, de

---

([1]) C'est l'analogie qui a guidé M. Potoček [1, p. 5], lequel a fait grand usage de déterminants analogues à ceux de Fredholm.

([2]) C'est l'analogie qui a servi de point de départ à M. Obrechkoff [1] dans un travail sur lequel nous reviendrons plus loin (Supplément à la Note C).

degré r au plus, qui est $\neq 0$ pour λ assez petit. On aura donc

$$(24) \qquad X_k(\lambda) = -\frac{\sum_j x_j D_{jk}(\lambda)}{D(\lambda)},$$

où $D_{jk}(\lambda)$ est le coefficient de $(\lambda a_{jk} - \delta_{jk})$ dans le développement de $D(\lambda)$. $D_{jk}(\lambda)$ est un polynome en λ de degré $r-1$. Le second membre n'est pas seulement holomorphe pour $\lambda = 0$, c'est encore une fonction méromorphe dans le plan complexe des λ et c'est même *une fraction rationnelle* en λ. Nous savons déjà que, dans le cas borné, $X_k(\lambda)$ est égal à cette fraction rationnelle et que, $X_k(\lambda)$ étant holomorphe pour $|\lambda| < 1$, les pôles de $X_k(\lambda)$ doivent être en module ≥ 1. Comme on a évidemment $D(\lambda) = \lambda^r \Delta\left(\frac{1}{\lambda}\right)$, nous vérifions une fois de plus que les racines de $\Delta(s)$ sont dans le cas borné, en module ≤ 1.

Qu'on soit dans le cas borné, ou non, le second membre de (24) est une fraction rationnelle en λ bien définie dont les pôles sont distincts de l'origine. Elle est donc développable en série entière en λ, convergente pour λ assez petit : $y_k + \sum_{n=1}^{\infty} y_k(n+1)\lambda^n$. Son origine montre qu'elle satisfait au système (23), que, par conséquent, les $y_k(n)$ satisfont au système (2) et comme, en faisant dans son expression, $\lambda = 0$, on vérifie que $y_k(1) = x_k(1)$, on en déduit $y_k(n) = x_k(n)$. Ainsi la série (22) qui représente $X_k(\lambda)$ est nécessairement convergente pour λ assez petit et sa somme est bien fournie par (24).

Les seuls pôles possibles de $X_k(\lambda)$ sont les racines de $D(\lambda)$, ils sont donc distincts de l'origine, et en nombre r. Les racines de $D(\lambda)$ sont d'ailleurs évidemment indépendantes de k et même de la solution particulière considérée des équations (2), déterminée par les $X_k(\lambda)$.

En particulier, quand $x_j = \delta_{jl}$, on a $x_j(n+1) = a_{jl}^{(n)}$, donc

$$(25) \qquad \delta_{kl} + \sum_{n=1}^{\infty} a_{kl}^{(n)} \lambda^n = -\frac{D_{lk}(\lambda)}{D(\lambda)}.$$

Expression des solutions $x_k(n)$ en fonction de n. — Considérons

d'abord le cas où le coefficient $D = \Delta(0)$ de λ^r dans $D(\lambda)$ est $\neq 0$. Alors $X_k(\lambda)$ est nul à l'infini et la décomposition de la fraction rationnelle $X_k(\lambda)$ en éléments simples ne comportera pas de partie entière.

En appelant c l'une quelconque des q racines distinctes $\lambda_1, \ldots, \lambda_q$ de $D(\lambda)$. $(q \leq r)$ et ρ l'ordre de multiplicité de cette racine c, $(\rho \leq r)$, on voit que $X_k(\lambda)$ est la somme de q expressions qu'on peut mettre, puisque c est $\neq 0$, sous la forme

$$T(\lambda) = \frac{B^{(1)}}{1-\frac{\lambda}{c}} + \frac{B^{(2)}}{\left(1-\frac{\lambda}{c}\right)^2} + \cdots + \frac{B^{(\rho)}}{\left(1-\frac{\lambda}{c}\right)^\rho}$$

$$= \sum_{n=0}^{\infty}\left[B^{(1)} + \cdots + \frac{B^{(\rho)}}{(\rho-1)!}(n+\rho-1)\cdots(n+1)\right]\frac{\lambda^n}{c^n}$$

$$= \sum_{n=0}^{\infty} Q(n)\left(\frac{\lambda}{c}\right)^n,$$

où $Q(n)$ est un polynome en n de degré $\rho - 1$

De sorte que $X_k(\lambda)$ est de la forme

$$(26) \qquad X_k(\lambda) = \sum_{n=0}^{\infty} \lambda^n \sum_{g=1}^{q} Q_{kg}(n) \frac{1}{(\lambda_g)^n}.$$

Il en résulte qu'en posant $\sigma_g = \frac{1}{\lambda_g}$, on aura l'expression générale des solutions du système (2), — qu'on soit ou non dans le cas borné, mais en supposant que zéro n'est pas racine de $\Delta(s)$ —, sous la forme

$$(27) \qquad x_k(n) = \sum_g (\sigma_g)^n R_{kg}(n),$$

où $R_{kg}(n) \equiv \lambda_g Q_{kg}(n-1)$ est un polynome en n de degré inférieur à l'ordre de multiplicité ρ_g de la racine σ_g de $\Delta(s)$.

C'est l'expression générale admise page 257 et qui se trouve établie, quand $\Delta(0) \neq 0$, pour toute valeur de l'entier n.

Dans le cas où $\Delta(s)$ a une racine nulle, alors si ρ_0 est son ordre de multiplicité, le développement de $X_k(\lambda)$ en fractions simples pourra comporter une partie entière, dont le degré sera au plus égal

NOTE B

GÉNÉRALISATION DES s_{jk}.

Partie principale de a_{jk}^{n}. — Dans le développement de la méthode directe, dont une partie a été exposée plus haut (p. 187), M. Doeblin [2, 4] a étendu heureusement au cas singulier l'usage de la partie principale de la probabilité P_{jk}^{n}. On peut opérer de même avec les quantités plus générales a_{jk}^{n} de la Note A dans le « cas borné ».

On a vu que

$$a_{jk}^{(n)} = \varpi_{jk}(n) + \varepsilon_{jk}(n),$$

où $\varpi_{jk}(n)$ est une fonction quasi-périodique de n et $\varepsilon_{jk}(n)$ tend vers zéro avec $\frac{1}{n}$. On peut appeler $\varpi_{jk}(n)$ une « partie principale » de $a_{jk}^{(n)}$. Mais, de plus, on a vu (p. 259), que, pour $n > r$,

$$\varepsilon_{jk}(n) = \sum_{g=1}^{g=q'} (\sigma_g)^n \, w_{jkg}(n) \qquad (q' \leq r),$$

où les $|\sigma_g| < 1$. Soit ρ un nombre réel compris entre 1 et les $|\sigma_g|$ qui sont tous < 1, on aura

$$\varepsilon_{jk}(n) = \rho^n \sum_{g} \left(\frac{\sigma_g}{\rho}\right)^n w_{jkg}(n),$$

et puisque les $w_{jkg}(n)$ sont des polynomes :

$$|\varepsilon_{jk}(n)| \leq \rho^n \, w'_{jk}(n),$$

où $w'_{jk}(n)$ tend vers zéro avec $\frac{1}{n}$.

Dès lors, $\varepsilon_{jk}(n)$ converge exponentiellement vers zéro, la série $\sum_{n=1}^{n=+\infty} \varepsilon_{jk}(n)$ est absolument convergente, et, par suite aussi, la

série

$$s'_{ik} = \sum_{n=1}^{n=+\infty} [a_{jk}^{(n)} - \varpi_{jk}(n)]$$

Généralisation de s_{ik}. — Ici s'_{ik} est une généralisation de s_{ik}, qui se réduit à s_{ik} dans le cas non oscillant où les a_{ik} sont des probabilités p_{ik} de Markoff. On peut envisager une autre généralisation.

On sait (Fréchet [152], p. 7), que dans le cas borné, $a_{jk}^{(n)}$ a une limite en moyenne Π_{jk}. On va étudier le comportement asymptotique de

$$R_{jk}^{(n)} = \sum_{t=1}^{t=n} (a_{jk}^{(t)} - \Pi_{jk}) = \sum_{t=1}^{t=n} [a_{jk}^{(t)} - \varpi_{jk}(t)] + \sum_{t=1}^{t=n} [\varpi_{jk}^{(t)} - \Pi_{jk}] = s'_n + s''_n.$$

Or (Fréchet [1], p. 6, 7),

$$\varpi_{jk}^{(t)} = \sum_{g=q''}^{g=r} \lambda_{kg} e^{ti\mu_g} v_{jg} + \Pi_{jk},$$

où les μ_g sont réels et non congrus à 2π et où $t > r$; donc,

$$(1) \qquad s''_n - s''_{r-1} = \sum_{t=r}^{t=n} [\varpi_{jk}^{(t)} - \Pi_{jk}] = \sum_g \lambda_{kg} v_{jg} \frac{e^{ri\mu_g}}{1 - e^{i\mu_g}} (1 - e^{(n-r+1)i\mu_g}),$$

et

$$\frac{s''_1 + \ldots + s''_n}{n} - \frac{s''_1 + \ldots + s''_{r-1}}{n} - \frac{n - r + 1}{n} s''_{r-1}$$
$$= \sum_g \lambda_{kg} v_{jg} \frac{e^{ri\mu_g}}{1 - e^{i\mu_g}} \left[\frac{n - r + 1}{n} - \frac{e^{i\mu_g}}{1 - e^{i\mu_g}} \left(\frac{1 - e^{(n-r+1)i\mu_g}}{n}\right)\right].$$

Dès lors, on voit que s''_n converge en moyenne vers une certaine limite

$$(2) \qquad s''_{jk} = s''_{r-1} + \sum_g \lambda_{kg} v_{jg} \frac{e^{ri\mu_g}}{1 - e^{i\mu_g}},$$

et, par suite, $R_{jk}^{(n)}$ converge en moyenne vers la somme de s'_{jk} et de s''_{jk}. Dans le cas non-oscillant, $R_{jk}^{(n)}$ converge au sens ordinaire, donc aussi en moyenne, vers s_{jk}; nous pouvons donc *étendre la définition de* s_{jk}

en posant dans le cas borné le plus général,

(3) $$s_{jk} = \text{limite en moyenne}_{n \succ \infty} \sum_{t=1}^{t=n} (a_{jk}^{(t)} - \text{H}_{jk})$$

On peut supprimer dans cette formule les mots « en moyenne » dans le cas non oscillant où les $a_{jk}^{(n)}$ sont convergents. Mais il est certain qu'on ne peut le faire dans le cas borné le plus général, ou même dans le cas le plus général des probabilités en chaîne. Par exemple, quand, pour $r = 2$, le tableau des p_{ik} est $\begin{Vmatrix} 0 & 1 \\ 1 & 0 \end{Vmatrix}$, on a

$$p_{ik}^{(n)} = \begin{cases} p_{ik} & \text{si } n \text{ est impair,} \\ 1 - p_{ik} & \text{si } n \text{ est pair,} \end{cases}$$

d'où $\text{H}_{ik} = \frac{1}{2}$,

$$\sum_{t=1}^{t=n} (p_{jk}^{(t)} - \text{H}_{jk}) = \begin{cases} 0 & \text{si } n \text{ est pair,} \\ p_{jk} - \frac{1}{2} \neq 0 & \text{si } n \text{ est impair} \end{cases}$$

L'égalité (3) invite à considérer l'expression

(4) $$\text{U}_{jl}(n) = \sum_{t=1}^{t=n} [a_{jl}^{(t)} - \text{H}_{jl}] - s_{jl}.$$

On a, en vertu de (4),

$$\text{U}_{jl}(n) - \sum_k a_{jk} \text{U}_{kl}(n-1) = a_{jl} - \text{H}_{jl} - s_{jl} + \sum a_{jk} s_{kl}.$$

La limite en moyenne du premier membre est nulle comme celle de $\text{U}_{jl}(n)$ et $\text{U}_{kl}(n-1)$. Le second membre étant indépendant de n doit donc aussi être nul. On voit que l'expression $\text{U}_{jl}(n)$, considérée pour l fixe, est une fonction de i et de n, qui est solution du système

(H) $$u_j(n) - \sum_k a_{jk} u_k(n-1) = 0,$$

cette solution $\text{U}_{jl}(n)$ étant, de plus, soumise à la condition de converger en moyenne vers zéro.

Comme on est dans « le cas borné », les solutions de ce système sont bornées ; donc $U_{jl}(n)$ reste borné quand n varie ([1]).

Dans le cas où les a_{jk} sont les p_{jk} du problème des chaînes, et plus généralement dans le cas où les $a_{jk}^{(n)}$ sont des fonctions asymptotiquement périodiques, de période asymptotique N, comme on a

$$U_{jl}(n)=\sum_i a_{ji}^{(n-1)}U_{il}(1),$$

on voit que les $U_{jl}(n)$ seront aussi des fonctions asymptotiquement périodiques, de période asymptotique N.

Calcul des s_{jk} — On vient de voir que dans le cas borné le plus général, on a

$$s_{jk}-(a_{jk}-\Pi_{jk})=\sum_i a_{ji}s_{ik}. \tag{5}$$

Un raisonnement analogue permettrait de remplacer le second membre par $\sum_i s_{ji}a_{ik}$

En procédant comme dans le cas régulier (p 45), mais en prenant des limites en moyenne, on verrait de même que

$$\sum_i \Pi_{ji}s_{ik}=0, \tag{6}$$

$$\sum_i s_{ji}\Pi_{ik}=0 \tag{7}$$

D'ailleurs, dans le cas des probabilités où $\sum_k \Pi_{ik}=1$, en faisant la somme des équations obtenues pour $k=1, 2, \ldots$, on trouverait aussi :

$$\sum_i s_{ji}=0. \tag{8}$$

([1]) Si l'on pose, par exemple, $a_n=(-1)^{n+1}\sqrt{n}$, on voit facilement que a_n converge en moyenne vers $a=0$, sans que a_n soit borné. Notons que $|a_n-a|=\sqrt{n}$ ne converge pas en moyenne vers zéro (ni d'ailleurs vers une limite finie)

Dans le cas semi-régulier : celui où les $\Pi_{jk} = \Pi_k$ sont indépendantes du premier indice sans être toutes nulles, les $s_{j1}, s_{j2}, \ldots, s_{jl}$ et les $\Pi_1, \Pi_2, \ldots, \Pi_l$ sont respectivement solutions des mêmes équations linéaires aux seconds membres près, et nous avons vu (p. 262), que le système relatif aux Π_k n'a qu'un seul système de solution. Il en est donc de même du système relatif aux $s_{j1}, \ldots, s_{jl}$, et ceci quel que soit j. Finalement, on voit que dans le cas semi-régulier, on peut calculer les s_{jk} sans itération par résolution de systèmes d'équations ayant chacun un seul système de solutions.

NOTE C.

COMPORTEMENT ASYMPTOTIQUE D'UN SYSTÈME NON HOMOGÈNE D'ÉQUATIONS LINÉAIRES AUX DIFFÉRENCES FINIES DU PREMIER ORDRE A COEFFICIENTS CONSTANTS (¹)

Considérons un système d'équations d'itération de la forme

$$\text{(S)} \qquad y_k(n) - \sum_{i=1}^{i=l} a_{ki} y_i(n-1) = f_k(n-1) \qquad (k = 1, \ldots, l)$$

On peut démontrer que si les $f_k(n)$ sont des polynomes de degré $\leqq m$, alors, dans le cas borné, les solutions $y_k(n)$ sont de la forme

$$(1) \qquad y_k(n) = Q_k(n) + \varepsilon_k(n),$$

où les $Q_k(n)$ sont des polynomes de degré $\leqq m+1$, et les $\varepsilon_k(n)$ sont des solutions convergeant en moyenne vers zéro du système homogène (II) associé à (S), soit

$$\text{(II)} \qquad x_k(n) - \sum_i a_{ki} x_i(n-1) = 0.$$

De plus, si

$$(2) \qquad f_k(n) = F_k n^m + F_k^{(1)} n^{m-1} + \ldots + F_k^{(m)},$$

(¹) Pour les démonstrations, *voir* Fréchet [194]. [Dans ce mémoire, les calculs qui conduisent aux formules (14) et (16) montrent que le coefficient $-\frac{1}{2}$ qui y figure doit être remplacé par $-\frac{3}{2}$].

on peut préciser les valeurs des premiers termes de $Q_k(n)$:

$$(3)\quad Q_k(n) = \frac{1}{m+1}\left\{\sum_i H_{ki} F_i\right\} n^{m+1} + \left\{F_k + \sum_i s_{ki} F_i - \frac{3}{2}\sum_i H_{ki} F_i + \frac{1}{m}\sum_i H_{ki} F_i^{(1)}\right\} n^m + \ldots$$

En comparant les parties principales des deux membres de (S), on voit que le plus haut degré des $Q_k(n)$ est égal ou inférieur d'une unité au plus haut degré m, des $f_k(n)$.

En particulier, si les f_k sont des constantes F_k, la solution générale de (S) est :

$$y_k(n) = n\sum_i H_{ki} F_i + F_k + \sum_i s_{ki} F_i + \sum_i H_{ki}\mu_i + \varphi_k^{(n)},$$

où les μ_i sont des constantes arbitraires et où les $\varphi_k^{(n)}$ forment un système arbitraire de solutions, convergeant en moyenne vers zéro, du système (H).

Pour que, dans ce cas, (S) ait des solutions bornées, il faut et il suffit que $\sum_i H_{ki} F_i = 0$. Dans ce cas, parmi les solutions bornées figurent des solutions constantes, ces solutions sont

$$y_k = F_k + \sum_i s_{ki} F_i + \sum_i H_{ki}\mu_i,$$

où les μ_i sont des constantes arbitraires.

Dans le cas semi-régulier $\sum_i H_{ki}\mu_i$ peut être remplacé par une constante arbitraire indépendante de k.

Dans le cas non oscillant, si $f_k(n)$ est la somme d'un polynome $F_k n^m + \ldots + F_k^{(m)}$ et d'une quantité $\eta_k(n)$ qui converge exponentiellement vers zéro, les solutions correspondantes sont encore de la forme (3) où le polynome $Q_k(n)$ est encore de la forme (2), mais, de plus, la quantité $\varepsilon_k(n)$, qui n'est peut-être plus solution de (H), converge exponentiellement vers zéro.

Dans le cas borné, si $\frac{1}{n^\alpha} f_k(n)$ est borné et $\alpha \geqq 0$, toute solution de (S) est de la forme $y_k(n) = n^{\alpha+1} G_k(n)$ où $G_k(n)$ est borné.

Dans le cas ou les $a_{kl}^{(n)}$ sont des fonctions asymptotiquement périodiques, dont la période asymptotique commune est N, si, en outre les $f_k(n)$ sont de la forme $f_k(n) = n^{\alpha} A_k(n)$, où $A_k(n)$ est aussi une fonction asymptotiquement périodique, de période asymptotique N et où $\alpha \geq 0$, alors toute solution de (S) est de la forme $y_k(n) = n^{\alpha+1} B_k(n)$, ou $B_k(n)$ est une fonction asymptotiquement périodique, de période asymptotique N. Et l'on a

$$\lim_{n\to\infty} \text{ en moy } B_k(n) = \frac{1}{\alpha+1} \sum_l H_{kl} \Big[\lim_{n\to\infty} \text{ en moy } A_l(n) \Big]$$

En particulier, si $a_{kl}^{(n)}$ et $f_k(n)$ ne different respectivement de deux fonctions de période N que par les termes généraux de deux series absolument convergentes, toute solution de (S) sera de la forme $y_k(n) = n\varpi_k(n) + H_k(n)$, ou $\varpi_k(n)$ est de période N et ou $H_k(n)$ est borné.

La démonstration de ces propriétés s'appuie essentiellement sur la formule évidente de résolution des équations (S) :

$$(4) \qquad y_k(n) = \sum_l a_{kl}^{(n-1)} y_l(1) + f_k(n-1) + \sum_l [a_{kl} f_l(n-2) + \dots + a_{kl}^{(n-2)} f_l(1)],$$

et, d'autre part, sur les lemmes de convergence suivants (Fréchet [194])

Lemmes de convergence. — Soient deux suites $a_1, \dots, a_n, \dots$ $u_1, \dots, u_n, \dots$.

Premier lemme. — *Si* $\lim_{n\to\infty} a_n = a$ *et* $\Sigma u_n = S$, *et si la série* Σu_n *converge absolument* [1], *on a*

$$\lim_{n\to\infty} (a_1 u_n + \dots + a_n u_1) = aS.$$

Deuxième lemme. — *Si la suite des* a_n *converge en moyenne*

---

(1) Si la convergence de Σu_n n'est pas absolue, le théorème peut être mis en défaut.

vers a, alors l'expression

$$\frac{1}{n^{\alpha+1}}(a_1 n^\alpha + \ldots + a_{n-t+1} t^\alpha + \ldots + a_{n-1} 2^\alpha + a_n)$$

converge au sens ordinaire vers $\frac{a}{1+\alpha}$ *pour toute valeur constante de α telle que $\alpha + 1$ soit positif.*

TROISIÈME LEMME. — *Si la série Σu_n converge en moyenne vers une somme S, alors l'expression*

$$\frac{1}{n^\alpha}(u_1 n^\alpha + \ldots + u_{n-t+1} t^\alpha + \ldots + u_{n-1} 2^\alpha + u_n)$$

converge au sens ordinaire vers S pour α positif.

Pour démontrer les deuxième et troisième lemmes, il est tout aussi facile d'en démontrer les généralisations suivantes :

DEUXIÈME LEMME GÉNÉRALISÉ. — *Soient $u_1^{(n)}, \ldots, u_n^{(n)}$ des nombres $\geqq 0$, non décroissants, tels que $n u_n^{(n)}$ — ou plus généralement $n|u_n^{(n)} - u_1^{(n)}|$ — soit borné, et dont la somme*

$$S_n = u_1^{(n)} + \ldots + u_n^{(n)}$$

converge vers S; si a_n converge en moyenne vers a, l'expression

$$a_1 u_n^{(n)} + \ldots + a_n u_1^{(n)}$$

converge au sens ordinaire vers a S.

TROISIÈME LEMME GÉNÉRALISÉ. — *Soient $a_1^{(n)}, \ldots, a_n^{(n)}$ des nombres positifs non décroissants ($a_p^{(n)} \leqq a_{p+1}^{(n)}$); si la suite de leurs différences successives est non décroissante ($a_p^{(n)} - a_{p-1}^{(n)} \leqq a_{p+1}^{(n)} - a_p^{(n)}$), si en outre $a_n^{(n)} - a_p^{(n)}$ peut être rendu aussi petit que l'on veut en prenant $\frac{p}{n}$ assez grand, enfin, si, de plus, $a_n^{(n)}$ tend vers une limite a, alors quand la série Σu_n converge en moyenne vers S, l'expression*

$$a_n^{(n)} u_1 + \ldots + a_1^{(n)} u_n$$

converge au sens ordinaire vers aS.

On peut aussi démontrer la proposition suivante, que nous citons sans en avoir besoin ici :

QUATRIÈME LEMME. — *Si a_n converge en moyenne vers a, si $\sum u_n$ converge au sens ordinaire vers S, alors, pour que $a_1 u_n + \dots + u_1 a_n$ converge en moyenne vers aS, il suffit que les nombres a_n soient bornés dans leur ensemble* (¹).

CINQUIÈME LEMME. — *Si a_n et b_n sont deux fonctions asymptotiquement périodiques de n qui convergent en moyenne vers a et b respectivement et qui ont la même période asymptotique N, alors l'expression $\rho_n = \frac{a_1 b_n + \dots + a_n b_1}{n}$ est asymptotiquement périodique avec une période asymptotique égale à N, et elle converge en moyenne vers ab.*

On a ainsi

$$\rho_n = \varpi_k(n) + \varepsilon_k(n),$$

où $\varpi_k(n)$ est de période N et où $\varepsilon_k(n)$ est infiniment petit avec $\frac{1}{n}$. Dans le cas qui nous a été utile page 980, où a_n et b_n diffèrent respectivement de deux fonctions de n de période N par les termes généraux de deux séries absolument convergentes, $n\varepsilon_k(n)$ est borné.

CINQUIÈME LEMME GÉNÉRALISÉ. — *Dans les hypothèses du cinquième lemme, l'expression*

$$\frac{a_1 b_n n^\alpha + a_2 b_{n-1}(n-1)^\alpha + \dots + a_n b_1}{n^{\alpha+1}},$$

où $\alpha \geq 0$, est aussi une fonction asymptotiquement périodique de n qui converge en moyenne vers $\frac{ab}{\alpha+1}$.

Supplément à la note C.

Autre méthode. — On peut étendre la méthode indiquée page 268, au cas des systèmes non homogènes (S).

(¹) M. Obrechkoff a bien voulu nous informer que les quatre lemmes ci-dessus peuvent être aussi déduits de théorèmes généraux de convergence dus à MM. J. Schur [1] et Toeplitz [1].

On pose d'abord formellement

$$(5) \qquad Y_j(\lambda) = y_j(1) + \sum_{n=1}^{\infty} \lambda^n y_i(n+1).$$

S'il existe un système de solutions de (S) tel que la série entière du second membre converge pour λ assez petit et $j = 1, \ldots, r$, et s'il en est de même pour la série

$$F_j(n) = \sum_{n=1}^{\infty} \lambda^n f_j(n),$$

les $Y_j(\lambda)$ seront solutions du système linéaire, en général non homogène,

$$(6) \qquad \sum_i (\lambda a_{ji} - \delta_{ji}) Y_i(\lambda) = -F_j(\lambda) - y_j$$

[en posant $y_j = y_j(1)$].

Dans le cas général, on peut résoudre, au moins pour λ assez petit, par la règle de Cramér, et l'on trouve

$$(7) \qquad Y_k(\lambda) = -\frac{\sum_j y_j D_{jk}(\lambda)}{D(\lambda)} - \frac{\sum_j D_{jk}(\lambda) F_j(\lambda)}{D(\lambda)}.$$

On voit, réciproquement comme page 270 que : *sous la seule condition que les fonctions données* $F_j(\lambda)$ *soient holomorphes à l'origine, il existe* un système de fonctions $Y_k(\lambda)$, holomorphes à l'origine, solutions du système (6) et, par conséquent, en vertu de (5), *un système de solutions* $y_i(n)$ *du système* (S) *d'équations d'itération, prenant les valeurs arbitrairement données* y_i *pour* $n = 1$

On aperçoit immédiatement que, selon le choix des fonctions $f_k(n)$, et par suite des fonctions $F_k(\lambda)$, une grande variété de formes et de natures des solutions $y_i(n)$ pourra être obtenue. En particulier, la connaissance de la nature des singularités des fonctions $F_k(\lambda)$ permettra d'apporter des précisions sur le comportement asymptotique des $y_i(n)$. C'est cette méthode qui a été employée par

M. Obrechkoff pour étendre et compléter les résultats de notre Mémoire résumé au début de la Note C. Nous renvoyons pour ces compléments à une Note aux *Comptes rendus* de M. Obrechkoff [1] et à l'article (en cours d'impression) développant cette Note.

On observera que la solution (7) se présente sous la forme d'une somme de deux termes, l'un correspondant aux solutions $x_i(n)$ du système homogène (H) qui prennent les valeurs données $y_i(1)$ pour $n = 1$, l'autre correspondant à la solution particulière de (S) se réduisant à des valeurs nulles pour $n = 1$.

NOTE D

LES RACINES RATIONNELLES DE L'ÉQUATION SÉCULAIRE APPLIQUÉE A UN PROBLÈME DE PROBABILITÉ

Peut-être quelque algébriste s'intéressera-t-il à la question 2° posée à la fin de cette Note en termes purement algébriques, mais qui se trouve suggérée par les considérations arithmétiques suivantes, appliquées à un problème de probabilité?

On a vu que dans le cas régulier les racines de l'équation en s $\Delta(s) = 0$ sont en module au plus égales à l'unité, que la seule racine de module 1 est l'unité et que l'unité est racine simple. Ces propriétés se vérifieraient en particulier dans le problème du mélange des urnes de la page 12.

Propriété arithmétique. — *Dans le cas symétrique de ce problème*, en appelant ainsi le cas où $\alpha = \nu$ et $B = N$, M. Hostinský [18] a observé en outre cette curieuse propriété arithmétique, que *les racines de l'équation en s sont réelles et rationnelles pour les premières valeurs de l'ordre du déterminant* $\Delta(s)$.

Plus précisément il a constaté cette propriété pour tous les ordres $\leqq 6$. En appelant $\Delta_r(s)$ le déterminant $\Delta(s)$ d'ordre $r+1$, il a

obtenu les expressions suivantes (¹)

$$\Delta_2(s) = -(s-1)\,s\left(s+\frac{1}{2}\right),$$

$$\Delta_3(s) = (s-1)\left(s-\frac{1}{3}\right)\left(s+\frac{1}{9}\right)\left(s+\frac{1}{3}\right),$$

$$\Delta_4(s) = -(s-1)\left(s-\frac{1}{2}\right)\left(s-\frac{1}{8}\right)\left(s+\frac{1}{8}\right)\left(s+\frac{1}{4}\right),$$

$$\Delta_5(s) = (s-1)\left(s-\frac{3}{5}\right)\left(s-\frac{7}{25}\right)\left(s-\frac{1}{25}\right)\left(s+\frac{3}{25}\right)\left(s+\frac{1}{5}\right),$$

$$\Delta_6(s) = (s-1)\left(s-\frac{2}{3}\right)\left(s-\frac{7}{18}\right)\left(s-\frac{1}{6}\right)s\left(s+\frac{1}{9}\right)\left(s+\frac{1}{6}\right).$$

Enfin, nous apprenons, en dernière heure, que M. Rawles [2] a établi, indépendamment de MM. Hostinský et Potoček (¹), la validité pour tout entier r des propriétés constatées par ceux-ci pour $r \leqq 6$ Il a, en effet, pu développer directement le déterminant $\Delta(s)$ sous la forme

$$\Delta(s) = (-1)^{r+1}\prod_{l=0}^{r}\left\{s-\frac{(r-l)^2-l}{r^2}\right\}$$

Nous avons eu la curiosité de chercher si cette propriété se conserve dans le cas non symétrique du mélange des urnes. Les formules étant alors beaucoup plus compliquées, nous nous sommes d'abord contenté de vérifier que *la propriété signalée par M. Hostinský subsiste dans ce cas plus général pour les ordres* 3 *et* 4 (*voir* p. 288 pour un ordre quelconque) Nous avons obtenu

$$\Delta_3(s) = -(s-1)\left(s-1+\frac{u+v}{uv}\right)\left(s-1+2\,\frac{u+v-1}{uv}\right),$$

$$\Delta_4(s) = (s-1)\left(s-1+\frac{u+v}{uv}\right)\left(s-1-\frac{6}{uv}+3\,\frac{u+v}{uv}\right)\left(s-1-\frac{2}{uv}+2\,\frac{u+v}{uv}\right).$$

(¹) Plus récemment, M. Potoček a fait observer que toutes les racines de $\Delta_r(s) = 0$ peuvent être exprimées dans les cas considérés par M. Hostinský, c'est-à-dire pour $r \leqq 6$, par la formule $s_l = \dfrac{(r-l)^2-l}{r^2}$, où $l = 0, 1, \ldots, r$, formule qu'on vérifiera facilement dans ces cas.

Ces formules redonnent, quand $u = v = B = r$, les formules obtenues par M. Hostinský pour Δ_2 et Δ_3 ; elles prouvent, en outre que si u, v, B, r sont des entiers, les racines de $\Delta_2(s)$ et $\Delta_3(s)$ sont encore des nombres réels et rationnels.

Propriété algebrique. — La propriété arithmetique obtenue peut se mettre sous une forme algébrique plus génerale.

En posant, pour simplifier l'écriture, $p_k = p_{k,k+1}$, $q_k = p_{k,k-1}$, le déterminant $\Delta_r(s)$ prend, dans le probleme du mélange des urnes [où sont vérifiées les formules (4) ci-après], la forme

$$\Delta_r(s) = \begin{vmatrix} 1-p_0-s & q_1 & 0 & 0 & \cdots & & 0 \\ p_0 & 1-p_1-q_1-s & q_2 & & & & \\ 0 & p_1 & 1-p_2-q_2-s & & & & \\ 0 & 0 & p_2 & \cdot & & & \\ & \cdot & 0 & \cdot & & & \cdot \\ \cdot & \cdot & \cdots & \cdot & & 0 & 0 \\ & \cdot & \cdot & \cdots & \cdot & q_{r-1} & 0 \\ & & \cdot & \cdot & \cdot\ 1-p_{r-1}-q_{r-1}-s & & q_r \\ 0 & & & \cdot & 0\quad 0 & p_{r-1} & 1-q_r-s \end{vmatrix}$$

On peut alors affirmer que : *si les nombres $p_0, p_1, \ldots, q_1, \ldots, q_r$, rationnels ou non, réels ou non*, sont tels que l'on ait, pour $r = 2$,

$$(2) \qquad p_0 - q_2 = 2(p_1 - q_1),$$

ou pour $r = 3$

$$(3) \quad p_0 = 3(p_1 - p_2), \qquad q_3 = 3(q_2 - q_1) \qquad \text{et} \qquad p_1 - q_2 = 2(p_2 - q_1),$$

alors, dans le premier cas, les racines de $\Delta_2(s)$, dans le second cas celles de $\Delta_3(s)$, sont *des polynomes* par rapport aux p et aux q.

On trouve que :

sous la condition (2), les racines de $\Delta_2(s)$ sont

$$s_0 = 1, \qquad s_1 = 1 - \frac{p_0 + q_2}{2}, \qquad s_2 = 1 - (p_0 + 2q_1) = 1 - (q_2 + 2p_1),$$

sous les conditions (3), les racines de $\Delta_3(s)$ sont

$$s_0 = 1, \quad s_1 = 1 - \frac{p_0 + q_3}{3}, \quad s_2 = 1 - p_1 - \frac{2q_3}{3}, \quad s_3 = 1 - q_3 - 3p_2.$$

Il est inutile de reproduire ici les calculs qui nous ont fourni ces résultats. Ceux-ci une fois formulés, il suffit, pour les démontrer, de former le polynome ayant ces quantités pour racines, et de constater son identité avec $-\Delta_2(s)$ ou $\Delta_3(s)$ respectivement.

On s'assure aussi facilement que les conditions (2) et (3) sont satisfaites dans le cas du mélange des urnes. Et, dans ce cas, les p et q sont des nombres réels et rationnels. Mais c'est là un cas particulier de la proposition ci-dessus qui subsiste quand les p et q ne sont pas des nombres rationnels.

Tout cela suggere la question arithmétique et la question algébrique suivantes :

Soit le déterminant (1)

1° quand les p_k, q_k sont donnés par les formules de la page 15

$$p_k = \frac{(u-k)(B-k)}{uv}, \qquad q_k = \frac{k(v-B+k)}{uv}, \tag{4}$$

où u, v, B sont des entiers quelconques et où, si, par exemple, $v \geqq B$, r est le plus petit des entiers u et B, déterminer les valeurs de r pour lesquelles les racines de $\Delta_r(s)$ sont des nombres réels et rationnels. (Nous avons vu que $r = 2$ et 3 sont des solutions) ([1]).

2° quand les p_l, q_l sont des nombres réels ou complexes arbitraires, déterminer les relations qui doivent exister entre les p_l, q_l pour que les racines de $\Delta_r(s) = 0$ soient des polynomes en p_l, q_l. On a vu que cela est possible pour $r = 2$ et 3 et que les relations (2) et (3) sont respectivement des relations suffisantes.

([1]) Nous avons pu, pendant la correction des épreuves, résoudre complètement ce problème en prouvant que toutes les valeurs entières de r sont aussi des solutions (*Voir* le Supplément à la Note D, page suivante). Par contre, la question 2° reste ouverte.

Supplement à la Note D

Nous allons calculer $\Delta(s)$ dans le cas du mélange des urnes en étendant avec quelques modifications de forme, à des valeurs entieres quelconques de u, v, B, N, la méthode employée par M. Rawles [1, 2] dans le cas ou $u = v = B = N = r$.

Les racines de $\Delta(s) = 0$ sont les valeurs de s pour lesquelles existe un systeme de solutions non toutes nulles de

$$s x_k = \sum_j p_{kj} x_j \qquad (k = 0, 1, 2, \ldots, r),$$

ou, d'apres les formules (4) de la page 287, de

$$\frac{k(v-B+k)}{uv} x_{k-1} + \left[1 - s - \frac{k(v-B+k)}{uv} - \frac{(k-u)(k-B)}{uv}\right] x_k + \frac{(k-u)(k-B)}{uv} x_{k+1} = 0,$$

$$(5) \quad k(k+v-B) x_{k-1} - [k(k+v-B) + (k-u)(k-B) + (s-1)uv] x_k + (k-u)(k-B) x_{k+1} = 0$$

Les coefficients de ces équations récurrentes étant des polynomes par rapport a l'entier k, on peut employer comme M. Rawles la methode classique qui consiste à essayer de trouver des solutions de ces équations qui soient de la forme

$$x_k = A_0 + A_1 k + A_2 k(k-1) + \ldots + A_k k^r$$

qu'on peut écrire

$$(6) \qquad x_k = \sum_{n=0}^{r} A_n k^{(n)},$$

en posant $k^{(0)} = 1$ et $k^{(n)} = k(k-1)\ldots(k-n+1)$. Pour transformer l'équation (5) de façon que le premier membre devienne une combinaison linéaire des $k^{(n)}$, on commence par transformer les coefficients des x, en écrivant (5) sous la forme

$$[k(k+1) + k(v-B-1)] x_{k-1}$$
$$- [2(k+1)(k+2) + (k+1)(v-u-2B-6)$$
$$+ u + 2B - v + 2 + uB + (s-1)uv] x_k$$
$$+ [(k+2)(k+3) - (k+2)(u+B+5) + (u+2)(B+2)] x_{k+1} = 0.$$

Ceci devient, quand on y substitue (6), $\Sigma A_n \alpha_n = 0$, avec

$$\begin{aligned}\alpha_n = (k+1)^{(n+2)} + k^{(n+1)}(v - B - 1) \\ - \{2(k+2)^{(n+2)} + (k+1)^{(n+1)}(v - u - 2B - 6) \\ + k^{(n)}[u + 2B - v + 2 + uB + (s-1)uv]\} \\ + (k+3)^{(n+2)} - (k+2)^{(n+1)}(u + B + 5) + (k+1)^{(n)}(u+2)(B+2).\end{aligned}$$

On peut exprimer α_n comme combinaison linéaire de $k^{(n+2)}$, $k^{(n+1)}$, $k^{(n)}$, $k^{(n-1)}$, au moyen de la formule facile à établir

$$(k+1)^{(m)} = k^{(m)} + mk^{(m-1)}$$

et de celles qu'on en déduit

$$(k+2)^{(m)} = k^{(m)} + 2mk^{(m-1)} + m(m-1)k^{(m-1)},$$

$$(k+3)^{(m)} = k^{(m)} + 3mk^{(m-1)} + 3m(m-1)k^{(m-2)} + m(m-1)(m-2)k^{(m-3)}.$$

En les utilisant dans α_n, les termes en $k^{(n+2)}$, $k^{(n+1)}$ disparaissent et il reste

$$\alpha_n = k^{(n)}[(n-u)(n-v) - n - suv] + k^{(n-1)}n[(n-u-1)(n-B-1)].$$

En résumé, l'équation (5) devient

$$\begin{aligned}\sum_{n=0}^{r} A_n \{k^{(n)}[(n-u)(n-v) - n - suv] \\ + k^{(n-1)}n[(n-u-1)(n-B-1)]\} = 0\end{aligned}$$

ou

$$\begin{aligned}\sum_{n=0}^{r} k^{(n)}\{A_n[(n-u)(n-v) - n - suv] \\ + A_{n+1}(n+1)[(n-u)(n-B)\} = 0\end{aligned}$$

Il suffit alors, pour vérifier cette équation, de prendre

$$(7) \quad A_n[(n-u)(n-v) - n - suv] + A_{n+1}(n+1)[n-u)(n-B)] = 0,$$

pour $n = 0, 1, \ldots, r$. En particulier, la dernière équation se réduit à

$$A_n[(r-u)(r-v) - r - suv] = 0,$$

puisque, r étant égal à u ou B, le coefficient de A_{r+1} est nul. On a donc : ou bien $A_r \neq 0$, et alors il faut prendre pour s la racine s_r obtenue en annulant pour $n = r$ l'équation

$$(8) \qquad (n-u)(n-v) - n - suv = 0;$$

ou bien $A_r = 0$ et l'équation obtenue en faisant $n = r-1$ dans (7) se réduit à

$$\{A_n[(n-u)(n-v)-nsuv]\}_{n=r-1} = 0.$$

Alors, ou bien $A_{r-1} \neq 0$ et s doit être égal à la racine s_{r-1}, etc. Finalement, la solution ne peut être obtenue par cette méthode que si l'on prend pour s l'une des valeurs $s_0, s_1, \ldots, s_r$, avec

$$s_l = \frac{(u-l)(v-l)-l}{uv}.$$

D'ailleurs, cela suffit, car quand on prend $s = s_l$, c'est après avoir pris $A_r = A_{r-1} = \ldots = A_{l+1} = 0$. Il reste alors à satisfaire (7) pour $n = 0, 1, 2, \ldots, l-1$. En particulier, pour $n = 0$,

$$A_0 uv(1-s) + A_1 uB = 0.$$

Il suffira de prendre $A_0 = 1$, puis de tirer successivement de ces l équations $A_1, A_2, \ldots, A_l$, ce qui sera possible puisque le coefficient de A_{n+1} est différent de zéro pour $n < r$. Les équations (7) sont donc bien satisfaites et les s_l sont bien des racines de $\Delta(s) = 0$.

Observons d'ailleurs que ce sont toutes les racines. Comme elles sont en nombre $r+1$, égal au degré de $\Delta(s)$, il suffit, en effet, de vérifier qu'elles sont distinctes ; or

$$uv(s_l - s_{l'}) = (l-l')[l+l'-(u+v+1)],$$

et si $l \neq l'$, on a

$$l+l' < 2r \leq u+B \leq u+v < u+v+1.$$

En résumé, *on peut décomposer explicitement $\Delta_r(s)$ sous la forme*

$$(9) \qquad \Delta_r(s) = (-1)^{r+1} \prod_{l=0}^{r} \left[s - \frac{(u-l)(v-l)-l}{uv} \right].$$

On observera que $\Delta_r(s)$ est déterminé connaissant le nombre de boules de chaque urne, indépendamment du nombre des boules de chaque couleur.

Nous voyons que nous avons résolu la première question de la page 287 : *dans le problème du mélange des urnes, les racines de l'équation « en s » sont toutes réelles et rationnelles*, non seulement

quel que soit le nombre total des boules, mais aussi quelle que soit leur répartition dans les urnes ou suivant la couleur.

[Après avoir envoyé à l'impression ce résultat, nous avons reçu de M. Rawles une lettre nous informant qu'il venait d'obtenir la même formule (9)].

Dans le cas particulier où $u = v = B = N$, on retrouve la décomposition en facteurs de $\Delta(s)$, publiée par M. Rawles et citée page 285.

Il est d'ailleurs évident que la méthode employée ci-dessus pourrait s'étendre à des cas plus généraux où les p_k, q_k seraient encore des fonctions polynomiales de k.

Par contre, la question 2° de la page 287 ne paraît pas relever de la même méthode et reste ouverte.

2° NOTES DIVERSES

NOTE E

SUR LES SOLUTIONS D'UN SYSTÈME CANONIQUE D'ÉQUATIONS DIFFÉRENTIELLES LINÉAIRES

Coefficients continus. Considérons le système

$$(E) \qquad \frac{dx_k}{dt} = \sum_{i=1}^{r} x_i\, U_{ik}(t) + F_k(t), \qquad (k = 1, \ldots, r)$$

La méthode des approximations successives sous sa forme habituelle s'applique au cas où les $U_{ik}(t)$ et $F_k(t)$ sont continus sur un segment S. Alors, s désignant un point de ce segment, la méthode consiste à former la suite des fonctions $x_{kn}(t)$ telles que

$$x_{kn}(t) = x_k(s) + \sum_{i=1}^{r} \int_s^t x_{i,n-1}(\tau)\, U_{ik}(\tau)\, d\tau + \int_s^t F_k(\tau)\, d\tau,$$

et à prouver que cette suite converge normalement (déf., en note, p. 215) sur S vers une fonction $X_k(t)$, de sorte qu'en passant à la limite, les

fonctions limites X_k vérifient les équations intégrales

$$(I)\quad x_k(t) = x_k(s) + \sum_{i=1}^{r} \int_s^t x_i(\tau)\, U_{ik}(\tau)\, d\tau + \int_s^t F_k(\tau)\, d\tau, \qquad (k = 1, \ldots, r),$$

et, par suite, les équations différentielles (E). En prenant $x_{k0}(t)$ identique à une solution, on voit de plus que cette solution est unique.

Quand on détermine explicitement les termes des séries fournies par cette méthode pour représenter les solutions, on constate, avec M. Hostinský [17], que la solution unique mentionnée ci-dessus est de la forme

$$(1)\qquad x_k(t) = \sum_j \left\{ x_j(s)\, \varphi_{jk}(s, t) + \int_s^t \varphi_{jk}(\tau, t)\, F_j(\tau)\, d\tau \right\},$$

où

$$(2)\quad \varphi_{jk}(s, t) = \delta_{jk} + \int_s^t U_{jk}(\tau_1)\, d\tau_1 + \sum_i \int_s^t \int_s^{\tau_1} U_{ji}(\tau_1)\, U_{ik}(\tau_2)\, d\tau_2\, d\tau_1 +$$
$$+ \sum_{i_1, \ldots, i_n} \int_s^t \int_s^{\tau_n} \cdots \int_s^{\tau_2} U_{ji_1}(\tau_1)\, U_{i_1 i_2}(\tau_2) \ldots U_{i_n k}(\tau_n)\, d\tau_n\, d\tau_{n-1} \ldots d\tau_1 + \ldots$$

et que ce développement est normalement convergent sur S. Les solutions obtenues sont non seulement dérivables mais à dérivées continues.

Remarque. — La méthode fournit d'abord la solution unique du système (I) obtenu par une intégration formelle de (E).

Quand les solutions ont chacune une dérivée continue, on a alors le droit d'en déduire par dérivation qu'elles vérifient (E). Si les U et F sont continues sur S, les dérivées des solutions seront nécessairement continues et la conclusion précédente sera légitime.

Quand les U et F ne sont pas continues sur S, on peut songer à substituer à l'intégration des équations différentielles (E), celle des équations intégrales (I) qui ne sont plus équivalentes aux équations (E).

Si les U et F sont mesurables et bornées, la méthode des approximations successives, sous sa forme usuelle, est encore applicable en ce qui concerne les équations (I).

Coefficients sommables. — Sous de légères complications de démonstration (*voir* Fréchet [199]), on peut l'étendre au cas où les U, F sans être nécessairement bornées sont sommables sur S : *si les $U_{ik}(t)$ et les $F_k(t)$ sont des fonctions sommables sur* S; 1° *il existe un système de solutions et un seul du système d'équations intégrales* (I), *prenant pour $t=s$* (de S) *des valeurs données*, 2° *ces solutions sont des fonctions continues* (et même *absolument continues*) *données par les formules* (1) et (2), les mêmes que celles qui ont été établies par M. Hostinský dans le cas où les U et F sont continus, la convergence normale des séries (2) subsiste sur S.

Observons d'ailleurs que si ces fonctions ne vérifient pas partout les équations différentielles (E), elles les vérifient au moins presque partout.

En posant

$$A_{ik}^{(0)}(t)=\int_s^t U_{ik}(\tau)\,d\tau, \qquad B_k^{(0)}(t)=\int_s^t F_k(\tau)\,d\tau,$$

les équations (I) peuvent s'écrire, en introduisant des intégrales de Stieltjes

$$x_k(t)=x_k(s)+\sum_{i=1}^{r}\int_s^t x_i(\tau)\,d\,A_{ik}^{(0)}(\tau)+B_k^{(0)}(t),$$

où les $A_{ik}^{(0)}$, $B_k^{(0)}$ sont des fonctions absolument continues. Ces dernières fonctions sont, en particulier, continues et à variation bornée sur S.

Solutions continues et à variation bornée. — On aura alors une nouvelle extension du problème, si l'on se propose d'intégrer un système d'équations de la forme

$$\text{(I)}\quad x_k(t)=x_k(s)+\sum_{i=1}^{r}\int_s^t x_i(\tau)\,d\,A_{ik}(\tau)+B_k(t) \qquad (k=1,\ldots,r),$$

où les A_{ik} et B_k sont des fonctions continues et à variation bornée sur S, choisies arbitrairement.

Des modifications dans le détail des démonstrations suffisent (Fréchet [199]) pour étendre aussi à ce cas la méthode des approximations successives :

Quand les $A_{ik}(t)$ et les $B_k(t)$ sont des fonctions continues et à

variation bornée sur S (fonctions qu'on pourra supposer, pour simplifier les formules, nulles pour un point s de S).

1° *il existe un système et un seul de solutions des équations* (F) *prenant en s des valeurs données;*

2° *ces solutions sont continues et à variation bornée sur S,*

3° *elles sont données par les formules*

$$x_k(t) = \sum_j \left\{ x_j(s)\,\varphi_{jk}(s, t) + \int_s^t \varphi_{jk}(\tau, t)\, d\mathrm{B}_j(\tau) \right\}$$

où

$$\varphi_{jk}(s, t) = \delta_{jk} + \mathrm{A}_{jk}(t) + \sum_i \int_s^t \mathrm{A}_{ji}(\tau)\, d\mathrm{A}_{ik}(\tau) + $$

$$+ \sum_{i_1, \dots, i_n} \int_s^t \int_s^{\tau_n} \cdots \int_s^{\tau_2} \mathrm{A}_{ji_1}(\tau_1)\, d\mathrm{A}_{i_1 i_2}(\tau_2) \dots d\mathrm{A}_{i_n k}(\tau_{n+1}) + \dots$$

cette série convergeant normalement sur S.

Enfin, on peut démontrer, (Fréchet [199]), que *dans les trois cas*, les $\varphi_{jk}(s, t)$ *vérifient le système* :

$$\Phi_{jk}(s, t) = \sum_{i=1}^{r} \Phi_{ji}(s, u)\,\Phi_{ik}(u, t),$$
$$\Phi_{jk}(s, s) = \delta_{jk}$$
$$(j, k = 1, \dots, r)$$

NOTE F.

SOLUTION LA PLUS GÉNÉRALE DE L'ÉQUATION FONCTIONNELLE.

$$(\mathrm{F}_1) \qquad \mathrm{D}(s, t) = \mathrm{D}(s, u)\,\mathrm{D}(u, t) \qquad (s < u < t).$$

Remarque. — On a obtenu plus haut, page 221, la solution *jamais nulle*, la plus générale de cette équation fonctionnelle, cette solution étant valable quand s, t varient à l'intérieur d'un intervalle ST fini ou non.

Passons maintenant à la recherche de la solution la plus générale (bornée ou non, mais partout finie) de l'équation (F_1) *quand on cesse de la supposer jamais nulle*. Il y a d'abord évidemment la solution $\equiv 0$ que nous écartons. Nous admettons donc qu'il existe au moins un couple s, t, $(s \leqq t)$ tel que $D(s, t) \neq 0$.

Un premier cas est celui où aucun de ces couples n'est formé de nombres distincts. C'est-à-dire que $D(s, t) = 0$ pour $s < t$ et égal à 0 ou 1 (et au moins une fois égal à 1) pour $s = t$. Ce sera nécessairement une solution discontinue. Inversement, *toute fonction définie de cette manière est évidemment une solution de* (F_1) quel que soit l'ensemble sur lequel $D(s, s) = 1$.

Problème reduit. — Il nous reste à chercher les solutions de (F_1) telles que $D(s, t)$ soit $\neq 0$ pour au moins un couple (s, t) avec $s < t$. Comme on a déja obtenu la solution générale quand $D(s, t)$ est toujours $\neq 0$, on doit supposer qu'il existe aussi un couple (s', t'), tel que $D(s', t') = 0$.

Ceci étant, soit u_0 un point intérieur à un segment s, t, $(s < u_0 < t)$ tel que $D(s, t) \neq 0$. Alors $D(s, u_0) \neq 0$ et $D(u_0, t) \neq 0$. Soient S la borne inférieure des s tels que $D(s, u_0) \neq 0$ et $s < u_0$, T la borne supérieure des t tels que $D(u_0, t) \neq 0$ et $u_0 < t$ (S peut être $-\infty$, T peut être $+\infty$), on aura $S < u_0 < T$. Il est clair que si (s, t) est un intervalle *intérieur* à (S, T), on aura $D(t, t) \neq 0$. Et si s est un point *intérieur* à (S, T), on aura non seulement aussi $D(s, s) \neq 0$, mais encore $D(s, s) = 1$. Appelons un intervalle tel que ST, *un intervalle rouge*.

Il est clair que s'il y a plusieurs intervalles rouges, ils ne peuvent chevaucher et n'ont, au plus, deux à deux, qu'une extrémité commune. Ils forment donc une suite finie ou dénombrable d'intervalles $J_k = (S_k, T_k)$.

D'après une remarque faite plus haut, il y a pour chaque J_k une fonction $A_k(s)$ finie et $\neq 0$ pour $S_k < s < T_k$, telle que

$$D(s, t) = \frac{A_k(t)}{A_k(s)}$$

pour $(S_k < s \leqq t < T_k)$.

Pour sortir de l'intérieur de J_k, posons encore $A_k(s) = 1$ pour $s < S_k$

et pour $s \geq T_k$, et enfin

$$\Lambda_k(S_k) = \begin{cases} 1 & [\text{si } D(S_k, u_k) = 0], \\ \dfrac{1}{D(S_k, u_k)} & [\text{dans le cas contraire}], \end{cases}$$

$$\Lambda_k(T_k) = \begin{cases} 1 & [\text{si } D(u_k, T_k) = 0], \\ D(u_k, T_k) & [\text{dans le cas contraire}] \end{cases}$$

[u_k étant le point intérieur à S_kT_k choisi déjà pour définir $\Lambda(s)$ à l'intérieur de cet intervalle, comme u_0 pour ST].

Alors $\Lambda_k(s)$ est une fonction partout finie et $\neq 0$ et le rapport

$$D_k(s, t) = \frac{\Lambda_k(t)}{\Lambda_k(s)}$$

est aussi partout fini et $\neq 0$. On a

$$D_k(s, t) = D(s, t) \quad \text{pour} \quad S_k \leq s \leq t \leq T_k$$

et

$$D_k(s, t) = 1 \quad \text{pour} \quad s \leq t \leq S_k \text{ ou } T_k \leq s \leq t \text{ ou } s \leq S_k < T_k \leq t.$$

Or, s appartient à deux J_k au plus, t appartient à deux J_k au plus. En mettant à part au plus quatre intervalles rouges quand s et t sont donnés, tout autre intervalle rouge S_kT_k est tel que

$$s < t < S_k < T_k,$$
$$S_k < T_k < s < t,$$
$$s < S_k < T_k < t,$$

et l'on a alors $D_k(s, t) = 1$.

Il en résulte que le produit fini ou infini

$$\Delta(s, t) = D_1(s, t) D_2(s, t) \ldots D_k(s, t) \ldots$$

ne comprend aucun terme nul et comprend au plus quatre termes $\neq 1$. C'est donc un produit fini ou convergent. De plus, si $D(s, t) \neq 0$, s, t appartiennent à l'un, J_l, des J_k et si $S_l \leq s \leq t \leq T_l$, on a

$$D_k(s, t) = \begin{cases} 1 & \text{si } k \neq l, \\ D(s, t) & \text{si } k = l. \end{cases}$$

Alors

$$D(s, t) = \Delta(s, t),$$

et l'on a $D(s, t) \neq 0$.

Enfin, $\Delta(s, t)$ est une solution de (F_1). Car, les A_k étant $\neq 0$, on peut écrire

$$\Delta(s, u)\Delta(u, t) = \Pi\left[\frac{A_k(u)}{A_k(s)}\right]\Pi\left[\frac{A_k(t)}{A_k(u)}\right] = \Pi\frac{A_k(t)}{A_k(s)} = \Delta(s, t)$$

Répresentation de $D(s, t)$ par le produit de deux solutions de natures différentes. — Posons alors

$$\delta(s, t) = \frac{D(s, t)}{\Delta(s, t)}.$$

Il est clair que $\delta(s, t)$ est bien déterminé et fini pour tout couple (s, t) $(s \leq t)$ et que c'est une solution de (F_1), comme Δ et D. Mais $\Delta(s, t)$ est une solution toujours $\neq 0$, alors que $\delta(s, t)$ est nul en même temps que $D(s, t)$

On a donc mis $D(s, t)$ sous la forme d'un produit de deux solutions de (F_1) de natures différentes :

$$D(s, t) = \Delta(s, t)\delta(s, t)$$

D'une part, $\Delta(s, t)$ étant toujours $\neq 0$ peut être mis sous la forme

$$\Delta(s, t) = \frac{U(t)}{U(s)},$$

où $U(t)$ est partout fini et $\neq 0$.

D'autre part, $\delta(s, t)$ présente les caractéristiques suivantes :

$\delta(s, t) = 1$ quand (s, t) est intérieur à un intervalle rouge;

$\delta(s, t) = 0$ quand s et t n'appartiennent pas à un même intervalle rouge.

Restent les cas où s, t appartiennent à un même intervalle rouge S_kT_k, mais sans lui être tous deux intérieurs. On a, par exemple,

$$S_k = s < t \leq T_k,$$

ou

$$S_k \leq s \leq t = T_k.$$

Considérons d'abord le cas où $S_k = s < t < T_k$. On a

$$\delta(S_k, t) = 0 \quad \text{si} \quad D(S_k, t) = 0.$$

Lorsque $D(S_k, t) \neq o$, alors

$$\Delta(S_k, t) = \prod_l D_l(S_k, t).$$

Tout intervalle J_l est sans point commun avec (S_k, t) ou identique à J_k ou identique à J_j, en désignant par J_j un intervalle rouge, s'il en existe, dont l'extrémité droite coïncide avec S_k. Alors tous les $D_l(S_k, t)$ autres que $D_k(S_k, t)$ et éventuellement $D_j(S_k, t)$ sont égaux à 1.

Calculons d'abord

$$D_k(S_k, t) = \frac{A_k(t)}{A_k(S_k)}.$$

Comme $D(S_k, t) \not\leq o$ et qu'on a

$$D(S_k, t) = D(S_k, u_k)\, D(u_k, t)$$

ou

$$D(S_k, u_k) = D(S_k, t)\, D(t, u_k),$$

avec $D(u_k, t) \not\leq o$ ou $D(t, u_k) \not\leq o$, on aura aussi $D(S_k, u_k) \neq o$ et, par suite,

$$A_k(S_k) = \frac{1}{D(S_k, u_k)}.$$

D'après la définition de $A_k(t)$, (p. 295), on a donc

$$D_k(S_k, t) = D(S_k, t),$$

que u_k t soit o ou o.

Alors, ou bien il n'existe pas d'intervalle rouge J_j contigu à J_k en S_k, et dans ce cas

$$\Delta(S_k, t) = D_k(S_k, t) = D(S_k, t),$$

ou bien on a

$$\Delta(S_k, t) = D(S_k, t)\, D_j(S_k, t).$$

Mais $D(u_j, u_k)$ est nécessairement nul, sans quoi la somme de J_j et de J_k serait un seul intervalle rouge. Donc

$$o = D(u_j, u_k) = D(u_j, S_k)\, D(S_k, u_k),$$

avec $D(S_k, u_k) \neq o$. Par suite,

$$D(u_j, T_j) = D(u_j, S_k) = o,$$

et alors

$$1 = A_j(T_j) = A_j(S_k), \qquad 1 = A_j(t) \qquad \text{d'où} \quad D_j(S_k, t) = 1,$$

d'où encore

$$\Delta(S_k, t) = D(S_k, t),$$

et, par suite,

$$\delta(S_k, t) = 1.$$

Ainsi, pour $S_k < t < T_k$, on a

$$\delta(S_k, t) = 1,$$

et de même

$$\delta(t, T_k) = 1$$

On a donc aussi

$$\delta(S_k, T_k) = \delta(S_k, t)\,\delta(t, T_k) = 1.$$

Les seules valeurs de $\delta(s, t)$ qui restent à examiner sont celles telles que $\delta(S_k, S_k)$, $\delta(T_k, T_k)$, qui sont égales à 0 ou 1.

Finalement, la fonction $\delta(s, t)$ ne peut prendre que les valeurs 0 ou 1.

En résumé, *toute solution (continue ou non) de l'équation fonctionnelle* (F_1), *peut se mettre sous la forme du produit*

$$D(s, t) = \Delta(s, t)\,\delta(s, t),$$

de deux solutions de (F_1), *l'une* $\Delta(s, t)$ *qui n'est jamais nulle et peut se mettre sous la forme*

$$\Delta(s, t) = \frac{U(t)}{U(s)},$$

où $U(s)$ *est une fonction partout finie et* $\neq 0$, *l'autre* $\delta(s, t)$, *qui ne prend que des valeurs* 0 *ou* 1.

Il est clair que $\delta(s, t)$ aurait pu être directement défini comme $= 0$ si $D(s, t) = 0$ et à 1 si $D(s, t) \neq 0$; que, d'autre part, on aurait pu définir directement $\Delta(s, t)$ comme $= D(s, t)$ si $D(s, t) \neq 0$, et comme $\neq 0$ si $D(s, t) = 0$. Mais si, de cette façon, $\delta(s, t)$ est entièrement déterminé, $\Delta(s, t)$ ne l'est pas, et ce que nous avons fait a consisté à déterminer les valeurs encore indéterminées de $\Delta(s, t)$ quand $D(s, t) = 0$, et cela de sorte que $\Delta(s, t)$ vérifie (F_1).

On a obtenu la solution la plus générale en ce sens que, réciproquement, si $\Delta(s, t)$ et $\delta(s, t)$ sont deux solutions arbitraires de (F_1), leur produit est naturellement solution de (F_1); et nous voyons qu'on peut supposer $\Delta(s, t)$ partout $\neq 0$ et $\delta(s, t)$ constamment égal à 0

ou 1 (mais non nécessairement constant). Reste à voir comment construire les solutions $\Delta(s, t)$, $\delta(s, t)$ de la façon la plus générale.

Nous connaissons déjà la forme la plus générale de $\Delta(s, t)$, soit

$$\Delta(s, t) = \frac{U(s)}{U(t)},$$

où $U(s)$ est une fonction *arbitraire* mais *partout finie et* $\neq$ o.

En ce qui concerne la construction de $\delta(s, t)$, on a vu qu'il existe un nombre fini ou dénombrable d'intervalles I_k ne chevauchant pas et à l'intérieur desquels $\delta(s, t) = 1$. On pourra prendre *arbitrairement* la suite d'intervalles I_k, pourvu qu'ils ne chevauchent pas, poser $\delta(s, t) = 1$ quand (s, t) est intérieur à l'un quelconque de ces intervalles et poser $\delta(s, t) = 0$ quand le segment (s, t) n'appartient pas entièrement à l'un de ces intervalles. Il reste à définir $\delta(s, t)$ quand s ou t ou s et t sont extrémités de l'un de ces intervalles. Si (S_k, T_k) est l'un d'eux, et si $S_k < s < T_k$, on prendra arbitrairement pour $\delta(S_k, s)$ la valeur o ou 1, mais cette valeur sera indépendante de s. De même on prendra arbitrairement pour $\delta(s, T_k)$ la valeur o ou la valeur 1, mais la même quel que soit s, $(S_k < s < T_k)$. Alors, la valeur de $\delta(S_k, T_k)$ prise égale à $\delta(S_k, s)\,\delta(s, T_k)$, $(S_k < s < T_k)$ sera déterminée indépendamment de s. Seulement si deux intervalles (S_j, T_j), (S_k, T_k) ont une extrémité commune $T_j = S_k$ et si $S_j < s < T_j = S_k < t < T_k$ on prendra

$$\delta(s, T_j)\,\delta(S_k, t) = \delta(s, t) = 0.$$

Enfin, les quantités $\delta(S_k, S_k)$ et $\delta(T_k, T_k)$ seront prises égales à 1 si $\delta(S_k, T_k) = 1$; dans le cas contraire, elles seront prises arbitrairement égales à o ou 1.

Nous avons pu former *la solution la plus générale* (continue ou non, jamais nulle ou non, mais partout finie) *de l'équation fonctionnelle* (F_1). Bien entendu, cette solution partout finie peut n'être pas bornée.

INDEX BIBLIOGRAPHIQUE

DES AUTEURS CITÉS.

L'ouvrage de M. Hostinský [16] cité plus loin, contient, aux pages 60-63, une liste bibliographique étendue. Pour abréger, nous n'avons fait figurer ci-après que les publications citées dans le présent volume et ne se trouvant pas dans la liste de M. Hostinský. Les numéros entre crochets dans notre index font suite, pour un même auteur, à ceux de la liste de M. Hostinský (¹).

Liste complémentaire.

S. Bernstein [5], *Sur les liaisons entre les grandeurs aléatoires* (*C. R. Congr. Int. Mat. Zurich*, vol. 1, 1932, p. 288-309).

H. Copeland [1], *A mixture theorem for non conservative mechanical systems* (*Bull. Amer. Math. Soc.*, vol. XLII, 12, 1936, p. 8[illegible]5).

H. Cramér [1], *Random variables and probability distributions*, Cambridge Tract nº 36, 1937, p. 121.

W. Doeblin [1], *Sur les chaînes discrètes de Markoff* [*C. R. Ac. Sc.*, t. 203, 1936, p. 24-26 (Errata p. 592)].

[2], *Le cas discontinu des probabilités en chaîne* (*Publ. Fac. Sc. Univ. Masaryk*, nº 236, 1937, p. 1-13).

[3], *Exposé de la théorie des chaînes simples constantes de Markoff à un nombre fini d'états* (*Revue Mathém. de l'Union Interbalk.*, t. 3, 1938).

[4], *Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples* (*Thèse*, Paris, ou *Bull. Math. Soc. Roumaine des Sciences*, Bucarest, 1937).

[5] *Sur l'équation matricielle* $\Lambda^{(t+s)} = |\Lambda^{(t)}\Lambda^{(s)}|$ *et ses applications aux probabilités en chaîne* (*Bull. Sc. Math.*, 1938).

G. Elfving [1], *Zur Theorie der Markoffschen Ketten* (*Acta Soc. Scient. Fennicæ*, t. II, 1937).

(¹) Pour mes propres publications j'ai toutefois conservé le numérotage chronologique que j'ai adopté une fois pour toutes dans ma Notice.

R. Fortet [1], *Sur l'itération des substitutions algébriques linéaires à une infinité de variables et ses applications au problème des probabilités en chaîne* (*Thèse* ou *Revista de Ciencias*, Lima, 1938).

A. Fouillade [1], *Sur une conception de la théorie des probabilités en chaîne* (*Bull. Sc. Math.*, t. LXI, p. 269-287 et 295-302, 1937). Voir aussi *Sur l'itération de certaines substitutions linéaires* (*Mémoires de l'Acad. Roy. de Belgique*, 1933-34.)

[2] *Recherches sur l'itération des substitutions linéaires* (*Thèse*, Poitiers, 1938, ou *Mémoires de la Société Royale des Sciences de Liège*, t. II, 1937).

M. Fréchet [147], *Compléments à la théorie des probabilités discontinues en chaîne* (*Annali Sc. Norm. Sup. Pisa*, vol. II, 1933, p. 131-164).

[150], *Solution continue la plus générale d'une équation fonctionnelle de la théorie des probabilités en chaîne* (*Bull. Soc. Math. France*, vol. LX, 1932, p. 242-280).

[153], *idem*, Supplément (*Bull. Soc. Math. France*, vol. LXI, 1933, p. 182-185).

[152], *Comportement asymptotique des solutions d'un système d'équations linéaires et homogènes aux différences finies du premier ordre à coefficients constants* (*Public. Fac. Sc. Univ. Masaryk*, Brno, n° 178, 1933, p. 3-24).

[181], *Sul caso positivamente regolare nel problema delle probabilità concatenate* (*Giorn. Ist. Ital. Attuari*, Anno VII, 1936, p. 28-36).

[188], *Recherches modernes sur la théorie des Probabilités*, Premier Livre : *Généralités sur les probabilités. Variables aléatoires*, avec une Note de Paul Lévy (Fasc. III du tome I du *Traité des probabilités* par Émile Borel et divers auteurs), xvi + 368 pages; Gauthier-Villars, 1936.

[195], *Sulla mescolanza delle palline e sulle legge limite delle probabilità* (*Giorn. Ist. Ital. Attuari*, Anno VIII, 1937, p. 14-28).

[194], *Étude du comportement asymptotique des solutions d'un système d'équations linéaires non homogènes aux différences du premier ordre à coefficients constants* (*Ann. Soc. Mat. Polonaise*, t. XVI, 1937, p. 1-22).

[200], *Existence, unicité et expressions des solutions d'un système canonique d'équations différentielles linéaires à coefficients discontinus* (sous presse).

[181], *Voir* Hadamard [6] et Fréchet.

[199], *Sur l'intégration d'un système canonique d'équations différentielles linéaires à coefficients discontinus* (*Proceed. Benares Math. Soc.*, sous presse).

[193], *Sur quelques notions fondamentales du Calcul des Probabilités*, dans le Volume édité à l'occasion du Jubilé de M. Romanovsky par l'Université de Tachkent (sous presse).

G. Frobenius [4], *Ueber Matrizen aus nicht negativen Elementen* (*Sitzungsber. Akad. Wissensch. Berlin*, 1908, 1912, p. 456-477).

J. HADAMARD [4], *Cours d'analyse*, chez Hermann, t. II, 1930.

[5], *Sur le battage des cartes et ses relations avec la Mécanique Statistique* (*Comptes rendus du Congrès international des Mathématiciens de Bologne*, vol. 5, p. 133-139, Bologne, 1928).

J. HADAMARD [6] et M. FRÉCHET, *Sur les probabilités discontinues des événements en chaîne* (*Zeitsch. f. angew. Math. u. Mech.*, Bd. 13, 1933, p. 92-97).

E. HOPF [1], *On causality, statistics and probability* (*Journ. of Math. and Phys.*, vol. XIII, 1934, p. 50-102).

[2], *Ueber die Bedeutung der willkürlichen Funktionen für die Wahrscheinlichkeitstheorie* (*Jahresbericht der Deutsch. Math. Ver.*, Bd. 46, 1936, p. 179-194).

B. HOSTINSKÝ [14 *bis*], *Sur une équation fonctionnelle de la théorie des probabilités* (*Public. Fac. Sc. Univ. Masaryk*, Brno, n° 156, 1932).

[15], *Obrácené Markovovy řetězy* (*en tchèque*) (*Rozpravy II, Třídy České Akademie, Ročník* XLV, číslo 6, 1935, p. 1-5).

[16], *Méthodes générales du Calcul des Probabilités*, (66 pages, Gauthier-Villars, 1931).

[17], *Sur une classe d'équations fonctionnelles* (*Journ. de Math.*, t. 16, 1937, p. 267-284).

[18], *Sur les probabilités relatives aux variables aléatoires liées entre elles. Applications diverses* (*Ann. Inst. Henri Poincaré*, t. VII, 1937, p. 69-119).

[19], *Équations fonctionnelles relatives aux probabilités continues en chaîne* (exposé d'Analyse générale (sous presse), Hermann, Paris).

[21] Voir V. VOLTERRA.

B. HOSTINSKÝ [20] et J. POTOČEK [1], *Chaînes de Markoff inverses* (*Bull. Intern. Ac. Sc. Bohême*, 1935, p. 64-67).

A. KOLMOGOROFF [2], *Zur Theorie der Markoffschen Ketten* (*Math. Ann.*, Bd II, 1935, p. 155-160).

[3], *Anfangsgründe der Theorie der Markoffschen-Ketten mit unendlich vielen möglichen Zuständen* (*Recueil Mathématique de Moscou*, t. 1 [43], 1936, p. 607-610).

[4], en russe (*Bull. de l'Université d'État de Moscou*, vol. 1, 1937).

M. KONEČNÝ [1], *Sur la théorie des chaînes de Markoff* (*Pub. Fac. Sc. Univ. Masaryk*, fasc. 147, 1931, p. 17).

P. LÉVY [2], *Théorie de l'addition des variables aléatoires* (fascic. I de la *Collection de Monographies sur le calcul des probabilités*, dirigée par M. E. Borel, VIII-330 pages, Gauthier-Villars, 1937).

M. LUBLIN [1], *Sur les systèmes linéaires aux différences finies du premier ordre à coefficients constants* (*Revue de Math. spéc.*, 1932-1933).

G. MIHOC [1], *Sur les propriétés générales des variables statistiques enchaînées* [*Thèse* (en roumain), 1934].

[2], *Sur les lois-limites des variables liées en chaîne* (*Bulet. Facult. de Ştiinţe din Cernăuţi*, vol. X, 1936, p. 1-26).

Voir aussi ONICESCU et MIHOC.

R. de Mises [1], *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik* (574 pages, *Deuticke, Wien*, 1931).

[2], *Note on deduced probability distributions* (*Bull. Americ. Math. Soc.*, vol. XLIV, 1938).

O. Obrechkoff [1], *Sur les solutions d'un système d'équations linéaires aux différences finies du premier ordre* (*C. R. Ac. Sc.*, t. 205, 1937, p. 961).

O. Onicescu et Mihoc [1], *Sopra le leggi limite delle probabilità* (*Giorn. Istituto Ital. Attuari*, Ann. VII, 1936, p. 54-69).

[2], *La dépendance statistique. Chaînes et familles de chaînes discontinues* (dans la Collection d'Exposés d'Analyse générale, chez Hermann, 41 pages, 1937).

[3], *Sur une généralisation de l'urne de Bernoulli* (*Bull. Math. et Phys. École Polytechnique*, Bucarest, 7e année, 1936-1937, p. 1-37).

[4], *L'allure asymptotique de la somme des variables d'une chaîne de Markoff discontinue* (*C. R. Ac. Sc.*, t. 205, 1937, p. 481-482).

J. Potocek [1], *Sur la dispersion dans la théorie des chaînes de Markoff* (*Publ. Fac. Sc. Univ. Masaryk*, no 154, 1932, p. 1-27).

J. Potocek [2], *voir* Hostinský [19].

Lord Rayleigh [1], *Philosophical Magazine*, t. 10, 1880, p. 73, t. 37, 1919, p. 321, *Scientific Papers*, vol. I, p. 491, vol. 5, p. 256.

T. R. Rawles [1], *Operational methods in Fissiparous Propagation* (*Human Biology*, vol. 8, 1936, p. 106-135).

[2], *A problem of Laplace* (*Report of third Annual conf. on economics and stat.*, Colorado Springs, 1937).

V. Romanovsky [3], *Recherches sur les chaînes de Markoff, Premier Mémoire* (*Acta Mat.*, t. 66, 1935, p. 147-251).

G. Schulz [1], *Ueber Markoffsche Ketten* (*Zeitschrift f. angew. Math. u. Mecha.*, Bd. 11, 1931, p. 444).

[2], *Ueber das Summenproblem bei Markoffschen Ketten* (*Congrès internat. des math.*, *Zurich*, 1932, vol. 3, p. 250-251).

[3] *Zur Theorie des Galtonschen Brettes* (*Zeitsch. f. Physik*, Bd. 92, 1934, p. 747-754).

[4], *Grenzwertsätze für die Wahrscheinlichkeiten verketteter Ereignisse* (*Deutsche Mathematik*, Jahrg. 1, 1936, p. 665-699).

I. Schur [1], *Ueber lineare Transformationen in der Theorie der unendlichen Reihen* (*Journ. de Crelle*, Bd. 151, 1921, p. 79-117).

O. Toeplitz [1], *Ueber allgemeine lineare Mittelbildungen* (*Prace Matem.*, t. 21, 1911, p. 79-117).

V. Volterra [1] et B. Hostinský, *Opérations infinitésimales linéaires. Applications aux équations différentielles et fonctionnelles* (Gauthier-Villars, 1937, 231 pages).

INDEX ALPHABÉTIQUE.

(Les numéros correspondent aux pages.)

Battage des cartes, 16, 40, 199

Chaine (de Markoff), 23
Chaine (simple, 22, constante, 22)
Chemin, 189.
Consequent, 188
Convergence en moyenne arithmétique (ou au sens de Cesaro), 71, 259.
Cycle, 87.

Decomposable, 169, 173, 189
Dispersion, 74.

Ecart quadratique moyen, 74, 76, 79
Epreuves (suite discrete et continue d'epreuves), 22
Equation en « s » (*ou seculaire*), 105, 258, 284
Ergodique (principe), 11, 198, presque ergodique, 198.

Fonctions arbitraires (méthode des), 6
Frequence, 72, 73.

Groupement (cyclique, 182, final, 176, indécomposable, 173, de passage, 176)

Hasard (explication du), XV, 7.
Hostinský (cas de), 33, (méthode de), 214.
Homogène (cas), 248

Indecomposable, 168, 173, 190.

Kolmogoroff (méthode de), 204

Laplace (cas de), 13.

Markoff (méthode de), 27; (problème ou cas de), 23.
Mélange des urnes, 12, 49, 69, 75, 93.
Mouvement circulaire, 98, 121, 122.
Moyenne (principe de), 25.

Non-oscillant (cas), 114

INDEX DES NOTATIONS.

	Pages
$a_{ij}(s)$	225
$b_{ki}(t)$	225
$d(s)$	226
f_{hk}^{n}	73
$g_{..}$	174
p_{hk}	24
q_{ik}	28
$q_{ik}^{(n)}$	57
s_{ik}	31, 263, 276
s'_{ik}	275
r_i	67
u	13
v	13
w_{hk}	141

	Pages
A_h^{n}	71
$A_h^{l}(n)$	147
$A_{ij}(s)$	227
$B_{ij}(t)$	227
B	13
C_B	12, 14
C_l	178
D(s)	226
D(s, t)	220
D, $D^{(1)}$, $D^{(n)}$ (tableaux)	26, 29
E_k (événement)	23
(F_r) (équation)	220
$F_{hk}^{(n,N)}$	72
G	174
$G_{ij}^{(n)}$	77
(I) (condition)	24, 202
(I_1) »	24
(L) »	202
(L') »	206
L_n	77
L_{kl}	147
M	68
$M_h^{(n)}$	71
M_h	132
N	13
P_k	26, 30
$P_{hk}^{(n)}$	24
P_{hk}	37
(P) (condition)	24, 202
(P_1) »	25
$P_{ik}(s, t)$	202
$Q_{ik}^{(m,n)}$	60
$Q_{ik}^{(n)}$	177
$R_{ik}^{(n)}$	48
S (système)	23
$S_{ij}^{(n)}$	78
(T) (condition)	24, 202
(T'_1) »	38
(T_1) »	24
T_n	136
T	50
U (urne)	12
V (urne)	12
W_h	136
$Y_h^{(n)}$	67

CHAPITRE II.

Nombre fini d'états possibles.

Section I.

Cas de Markoff : Suite discrète d'épreuves.

Pages
Hypothèse de base. . . . . . [illegible]

Première méthode.

Étude du cas régulier par le principe de moyenne. . . [illegible]

a. Problème du comportement des probabilités itérées P^{n}_{hk} quand n croît

I. Le cas régulier : condition nécessaire . . . 26
Conditions suffisantes. . . . . [illegible]
Exemples. 32
Cas de Hostinský . . . 33
Sa généralisation . . . 35
Remarques. . . . . 3[illegible]

II. Valeurs des probabilités limites dans le cas régulier . . . 37

α. Cas de la limite constante 38
Application au battage des cartes. 40
Autre position du problème. . . . . 4[illegible]

β. Valeurs non nécessairement égales des limites P_k des P^{n}_{jk}. . 43
Application au mélange des urnes. . . . . 4[illegible]
Composition la plus probable. . . . . [illegible]
Interprétation analytique de l'itération des probabilités au moyen des groupes de transformations. . . . . [illegible]

III. Probabilités absolues et probabilités inverses. . . . . [illegible]
Loi de probabilité initiale . . . . [illegible]
Probabilités en chaînes inverses. . . . . [illegible]
Suite d'épreuves illimitée dans les deux sens. . . . . 6[illegible]
Probabilités absolues . . . . 6[illegible]
Probabilités inverses. 64

b. Étude d'une variable aléatoire « en chaîne » dans le cas régulier... 67

Valeur moyenne 67
Remarques 68
Application au mélange des urnes 69
Moyennes arithmétiques 70
Fréquence moyenne . . . . 7[illegible]
Dispersion d'une variable aléatoire 74
Application au mélange des urnes . . . . 7[illegible]
Dispersion de la moyenne arithmétique des valeurs observées au cours de n épreuves 76
Remarque 80

Pages.

Détermination générale des cas où σ = 0 :

I. Procédé direct 81
II. Procédé basé sur une décomposition en carrés de σ^2 84
III. Une condition plus simple dans le cas positivement régulier. 86

Dispersion des fréquences. 89
Application au cas de deux états possibles 90
Retour au mélange des urnes . 93
Exemple de Lord Rayleigh modifié (mouvement circulaire) 98
Exemple du cas positivement régulier. 99
Déplacement moyen . 100
Dispersion . . 101

DEUXIÈME MÉTHODE

Étude du cas singulier et compléments à l'étude du cas régulier au moyen de l'expression des P_{jk}^{n} en fonction de n. . 102

Nécessité et avantages d'une nouvelle méthode 102
Extension hors du Calcul des Probabilités. . . 104
Propriétés des racines de module 1 de « l'équation en s » 105

a *Comportement des probabilités itérées* 109

Les probabilités P_{jk}^{n} convergent *toujours* en moyenne arithmétique, leurs divers modes de comportement asymptotique quand *n* croît 109
Calcul des probabilités limites. . 114
Calcul des s_{q} 116
Discussion dans le cas où il n'y a que deux états possibles . 117
Exemple du cas oscillant. 120
Retour au mouvement circulaire. . . . 121
Deux exemples du cas semi-régulier

I. Mouvement circulaire . 122
II. Schéma d'urnes. 125

b *Variables en chaîne* 130

Valeurs moyennes 132
Dispersion. 134
Cas non oscillant 137
Cas général. Premières propriétés 138
Valeurs moyennes et dispersion des fréquences. 141

c *Limite de la loi réduite de répartition des sommes de variables enchaînées*. 144

Rappel d'un cas simple 144
Cas des variables enchaînées. 144
Méthode de Schulz et Mihoc 145

Pages

A. *Comportement asymptotique des moments dans le cas général :*

Relation de récurrence entre moments 145
Parties principales des moments 146
Cas d'une variable aléatoire semi-régulière 149

B. *Comportement asymptotique des moments dans le cas régulier*

Les moments ont des développements limités 154
Moments d'ordre deux 155
Moments d'ordre trois 156
Moments d'ordre supérieur 156
Application à la convergence de Λ_h^n et de f_{hk}^n 159
Limite de la loi réduite de répartition. 160
Précisions sur les moments d'ordre supérieur . 162
Calcul de θ_r^{n+1} 164

d. Étude du cas positivement régulier. 167

Remarque . 167
Condition pour qu'un, au moins, des H_{jk} soit nul 167
Condition pour le cas positivement régulier... 169
Remarque 170
Une propriété des racines de module 1 de l'équation en s.. 170
Conséquences.... 172
Nouvelle forme de la condition pour le cas positivement régulier. 173

e. Répartition des états possibles en groupements indécomposables en une épreuve. 173

Interprétation des tableaux décomposables . 173
Décomposition en groupements indécomposables.. 175

f. Répartition des groupements finals en sous groupements cycliques. 177

Examen de D(s)...... 177
Cas semi-régulier positif 178
Sous groupements cycliques...... 181
Les sous groupements cycliques dans le cas général 184
Critère du cas non oscillant 185

TROISIÈME MÉTHODE.

Méthode directe :

Introduction...... 187
États conséquents...... 188
Condition pour que l'ensemble, G, des états possibles soit indécomposable. 189
Propriété des groupements indécomposables...... 190
Construction directe des groupements indécomposables...... 191
Probabilités de sortir des groupements de passage 191
Cas semi-régulier 194

Pages

Démonstration directe de la répartition d'un groupement indécomposable en sous-groupements cycliques. .. . 195
Critères des différents cas . . . 197
Application des critères 197
Principes ergodique, presque ergodique . 198
Cas du battage des cartes . . 199
Stabilité à la Poisson . . 200

SECTION II

Cas d'une suite continue d'épreuves

Position du problème. 201
Existence de solutions discontinues 202
Simple repérage du temps 203

PREMIÈRE MÉTHODE

Réduction à un système d'équations différentielles

Méthode de Kolmogoroff 204
Démonstration de la réciproque 207
Extension de la méthode précédente 210
Extension intermédiaire 213

DEUXIÈME ET TROISIÈME MÉTHODES

Solution sous forme d'une série d'intégrales multiples d'ordres croissant

Les deux méthodes de M. Hostinský . 214
Première extension. 217
Seconde extension.. 217

QUATRIÈME MÉTHODE

Solution en termes finis.

Caractère de la solution 219
Étude de l'équation fonctionnelle (F_1).. 220
Remarque sur la condition (L').. 220
Solution continue la plus générale de l'équation fonctionnelle (F_1).. . 221
Retour à l'équation fonctionnelle (F_1)... 223
Solution la plus générale de (F_1) quand le déterminant des $\varphi_{jk}(s, t)$ reste $\neq 0$...................... 224
Remarque.... 228
Indétermination de la représentation des $\varphi_{jk}(s, t)$.......... 228
Remarques.. 231
Solution continue de (F_n) et (L').............. 231
Comportement asymptotique des solutions continues......... 233
Limites généralisées.......... 235
Condition (T')................................ 236
Remarques sur les solutions non négatives........... 239

Pages

Solution continue la plus générale de (F_r) vérifiant les conditions (L'), (T'), (P') dans le cas où $r = 2$ 241
Cas de $r > 2$ 245

Cas homogène

Une solution très générale 248
La solution à dérivées continues la plus générale 249
Cas des probabilités 250
Cas de $r = 1$ ou 2 252

COMPLÉMENTS DE MATHÉMATIQUES PURES

1° Quatre notes sur les systèmes d'équations linéaires aux différences finies du premier ordre à coefficients constants

Note A. — *Comportement asymptotique des solutions dans le cas des systèmes homogènes* 256

Expression des solutions 256
Convergence en moyenne 259
Valeurs des limites généralisées des $a_{jk}^{(n)}$ 260
Cas symétrique 264
Cas des racines simples 264
Supplément à la Note A
Seconde méthode de résolution du système (1) d'équations aux différences 268
Expression des solutions $x_k(n)$ en fonction de n 270
Exemples d'exception 273

Note B. — *Généralisation des s_{ik}* 274

Partie principale de $a_{ik}^{(n)}$ 274
Généralisation de s_{ik} 275
Calcul des s_{ik} 277

Note C. — *Comportement asymptotique des solutions d'un système non homogène d'équations linéaires aux différences finies du premier ordre à coefficients constants* 278

Résultats 278
Lemmes de convergence 280
Supplément à la Note C : Autre méthode 282

Pages

NOTE D. *Les racines rationnelles de l'équation séculaire rencontrée dans un problème de probabilité.* . . . 284

Propriété arithmétique . . . 284
Propriété algébrique . . . 286
Supplément à la Note D. . . . 288

2° NOTES DIVERSES

NOTE E. *Sur les solutions d'un système canonique d'équations différentielles linéaires* . . . 291

Coefficients continus . . . 291
Coefficients sommables . . . 293
Solutions continues et à variation bornée . . . 293

NOTE F. *Solution la plus générale de l'équation fonctionnelle* $D(s, t) = D(s, u)\,D(u, t)$, $(s \leq u \leq t)$. . . . 294

Remarque . . . 294
Problème réduit . . . 295
Représentation de $D(s, t)$ par le produit de deux solutions de natures différentes . . . 297

INDEX BIBLIOGRAPHIQUE . . . 301

INDEX ALPHABÉTIQUE . . . 305

INDEX DES NOTATIONS . . . 307

TABLE DES MATIÈRES. . . . 309